



REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD

Year 2005

Name of Author GUERRERA, J. C.

**COPYRIGHT**

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

**COPYRIGHT DECLARATION**

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

**LOANS**

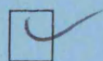
Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

**REPRODUCTION**

University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

- A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).
- B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.
- C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.
- D. 1989 onwards. Most theses may be copied.

***This thesis comes within category D.***



This copy has been deposited in the Library of UCL



This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.



QUANTITATIVE PHOSPHOPROTEOMICS  
OF PROTEIN ISOFORMS INVOLVED IN  
EGF SIGNALLING

*Ida Chiara Guerrera*

A thesis submitted to the University of London for the degree of  
Doctor of Philosophy

August 2005

University College London

UMI Number: U592859

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U592859

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346



## ABSTRACT

This thesis describes the development of a new, powerful concept and approach to study signalling pathways based on top-down quantitative phosphoproteomics. This approach, called the Phospho-SILAC strategy, combines a first step of in-house developed phosphoprotein enrichment using Immobilized Affinity Metal Chromatography (IMAC), and Stable Isotope Labelling in Cell Culture (SILAC) with labelled arginine containing six  $^{13}\text{C}$  atoms. NRK49F cells grown in medium containing  $^{13}\text{C}_6$  Arg (heavy arginine, hR) were stimulated for short times with EGF and the cell lysate obtained was mixed in a ratio 1:1 with the lysate from unlabelled cells, used as control. Phosphoproteins were then enriched from the protein mixture using IMAC and displayed on 2D gels, allowing multiple isoform separation. Individual post-translational changes in different proteins and in different isoforms of the same protein upon EGF stimulation are described.

Major focus was given to Hsp8 isoforms, as they were found to be differentially involved in EGF signalling when tested with antibodies against MAPK/CDK and PCK substrates. A potential MAPK phosphorylation site on Hsp8 N-terminal sequence was identified. The Phospho-SILAC strategy was also applied to Rho and Rab GTPases and their corresponding GDI (guanidine nucleotide dissociation inhibitor) molecules. Phosphorylation might regulate the binding between GTPase-GDI. The results obtained also suggested that RhoGDI-1 activity could be regulated by arginine methylation.

A potent new concept and methodology is outlined for wider investigations of molecular dynamics and for guiding the selection of candidate proteins and their modification for further studies.

# TABLE OF CONTENTS

ABSTRACT.....	2
TABLE OF CONTENTS .....	3
FIGURES.....	7
TABLES.....	9
ABBREVIATIONS.....	10
ACKNOWLEDGEMENTS.....	13
POSTERS AND PUBLICATIONS.....	14
<i>Posters and talks</i> .....	14
<i>Publications</i> .....	14
CHAPTER 1 .....	15
INTRODUCTION.....	15
<i>The role of proteomics in elucidating signalling networks</i> .....	15
Mass spectrometry in proteomics .....	16
Principles and instrumentation .....	16
Protein identification .....	17
Characterisation of post-translational modifications by mass spectrometry .....	24
Application of proteomics strategies in signalling studies .....	25
The bottom-up and top-down approaches in proteomics .....	25
Phosphoproteomics: the subproteome for signalling studies.....	27
Phosphoproteomics tools .....	28
Quantitative mass spectrometry.....	30
The Phospho-SILAC strategy .....	33
<i>EGF signalling</i> .....	34
EGF .....	35
EGF receptor.....	37
EGFR oligomerization and activation.....	38
EGFR transactivation .....	40
Signal pathways activated by the EGFR .....	41
Phospholipids metabolism: PLD, PLC, and PI3-K .....	43
MAPK (Mitogen activated protein kinase) pathway.....	44
The JAK and STAT pathways.....	49
GTPases in signalling.....	49
Chaperones involvement in EGF signalling .....	52
Signalling crosstalk .....	56
CHAPTER 2 .....	59
MATERIALS AND METHODS .....	59
<i>Materials</i> .....	59
<i>Cell Culture</i> .....	60
Culture of NRK49F.....	60

Freezing of NRK49F cells.....	60
EGF stimulation of NRK49F cells.....	61
Immunofluorescence Staining.....	61
In-vivo <sup>32</sup> P cells labelling.....	61
<i>IMAC purification of phosphoproteins</i> .....	62
<i>Protein assays</i> .....	63
Bicinchoninic acid protein assay.....	63
Modified Bradford protein assay .....	64
<i>Electrophoresis of proteins</i> .....	64
One-dimensional SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot.....	64
Two-dimensional- polyacrylamide gel electrophoresis (2D-PAGE).....	65
<i>Silver staining of polyacrylamide gels</i> .....	66
<i>In-gel tryptic protein digestion</i> .....	67
<i>Mass Spectrometric Analysis</i> .....	67
MALDI-TOF Mass Spectrometry.....	67
Mass Fingerprinting (MFP).....	68
ESI-Ion Trap Mass Spectrometry .....	68
CHAPTER 3 .....	70
PHOSPHOPROTEIN ENRICHMENT .....	70
<i>INTRODUCTION</i> .....	70
<i>RESULTS</i> .....	73
Development of phosphoprotein enrichment procedure .....	73
The phosphoproteome of NRK49F cells .....	76
Specificity of phosphoprotein capture .....	79
Comparison of IMAC-extracted proteins with the rat genome.....	88
<i>DISCUSSION</i> .....	94
CHAPTER 4 .....	98
FISHING FOR KINASE SUBSTRATES IN EGF SIGNALLING PATHWAYS .....	98
<i>INTRODUCTION</i> .....	98
<i>RESULTS</i> .....	101
Effects of EGF stimulation in NRK49F cells.....	101
Effects of EGF stimulation on specific substrates of MAPK, PKC and PKA .....	103
Phosphorylation kinetics.....	108
Structural proteins .....	108
Proteins involved in chaperone regulated redox network.....	109
Metabolic enzymes.....	115
<i>DISCUSSION</i> .....	117
EGF activates ERK1/2 in NRK49F cells .....	117
Phosphoproteome changes upon EGF stimulation .....	118
Kinase substrates immunodetection and identification .....	118
Limitations .....	120

Substrates identified are involved in chaperone-redox network .....	121
CHAPTER 5 .....	126
PHOSPHO-SILAC STRATEGY .....	126
<i>INTRODUCTION</i> .....	126
<i>RESULTS</i> .....	128
MALDI-TOF MS analysis of the <sup>13</sup> C <sub>6</sub> Arg labelled phosphoproteome .....	128
Labelling of proteins .....	128
Normalisation of the MALDI-TOF spectra .....	130
Data analysis .....	130
Quantitation.....	133
Protein kinase targets: Hsp8, Stress induced phosphoprotein 1 and Enolase-1 .....	134
Heat Shock Protein 8 (Hsp8).....	135
Stress induced phosphoprotein 1 .....	140
Enolase-1 .....	140
<i>DISCUSSION</i> .....	143
Hsp8 phosphorylation .....	143
Enolase-1 involvement in EGF signalling.....	147
CHAPTER 6 .....	150
PHOSPHO-SILAC STRATEGY WIDER APPLICATION IN SIGNALLING.....	150
<i>INTRODUCTION</i> .....	150
<i>RESULTS</i> .....	151
Other chaperones.....	152
Heat shock protein 4.....	153
Heat shock protein 5.....	154
Protein that act as metabolic enzymes .....	156
Protein acting in cell motility .....	157
Protein acting as signalling molecules.....	157
Collapsin response mediator protein .....	158
RhoGDI-1.....	159
Rab6 and Rab1.....	173
Proteins involved in protein synthesis .....	173
ATP synthase beta and RuvB-like protein 2.....	174
<i>DISCUSSION</i> .....	176
Proteins involved in EGF signalling, not recognised by kinase substrate antibodies.....	176
Different chaperones, same phosphorylation site.....	176
Proteins involved in protein synthesis.....	179
Signalling proteins.....	180
CHAPTER 7 .....	189
DISCUSSION .....	189
<i>Summary</i> .....	189
<i>The limitations of the Phospho-SILAC strategy</i> .....	190

<i>General observations on phosphorylation dynamics in the cell</i> .....	193
<i>Chaperones involved in EGF signalling</i> .....	194
<i>Rab/RabGDI vs. Rho/RhoGDI-1 interactions</i> .....	196
<i>Conclusions and outlook</i> .....	197
REFERENCES.....	199

# FIGURES

Figure 1. MALDI-TOF Mass Spectrometry .....	19
Figure 2. ESI/Ion-Trap Mass Spectrometry .....	22
Figure 3. Peptide fragmentation by ESI MS/MS.....	23
Figure 4. Activation of EGF receptor .....	40
Figure 5. Branching of the signal transduction pathways.....	42
Figure 6. Regulation of Ras/MAPK pathway by EGFR.....	46
Figure 7. MAPK phosphorylation systems.....	48
Figure 8. GTPases cycle.....	52
Figure 9. Scheme of Hsc70 involvement in Raf-1/ERK pathway and in PLC pathway.....	56
Figure 10. Scheme of different pathways involved in EGF signalling.....	58
Figure 11. Schematic representation of the protocol used for IMAC extraction of phosphoproteins .....	75
Figure 12. Silver stained 2D gels for total proteins, flow through, wash and phosphoproteins extracted from <sup>32</sup> P radio labelled NRK49F cells .....	77
Figure 13. Phosphoprotein enrichments specificity .....	78
Figure 14. Phosphoprotein fraction separated on 2D electrophoresis gel.....	80
Figure 15. Schematic description of the definitions used to classify acidic regions in protein sequences .....	91
Figure 16. Distribution of proteins of the rat genome.....	92
Figure 17. Comparison of the properties of isolated acidic motifs and acidic clusters for the experimentally identified proteins (black circles) with proteins in the rat genome (empty circles) .....	93
Figure 18. Analysis of enrichment of phosphoproteins after IMAC columns.....	102
Figure 19. Phosphoprotein fractions separated on 2D gel.....	103
Figure 20. Mapping of kinase substrates on 2D gel .....	106
Figure 21. Phospho-kinetics of Vimentin .....	109
Figure 22. Distribution of tryptic peptides in different isoforms of Hsp8 .....	110
Figure 23. Phospho-kinetics of Hsp8 .....	111
Figure 24. Phospho-kinetics of Stress induced phosphoprotein 1 .....	112
Figure 25. Phospho-kinetics of BiP .....	112
Figure 26. Phospho-kinetics of PDI .....	113



Figure 27. Phospho-kinetics of Glutathione S-transferase P .....	114
Figure 28. Phospho-kinetics of mitochondrial transmembrane GTPase .....	115
Figure 29. Phospho-kinetics of Enolase-1 .....	116
Figure 30. Phospho-kinetics of Phosphoglycerate mutase .....	116
Figure 31. Scheme of the Phospho-SILAC strategy .....	129
Figure 32. Example of selection of relevant attributions of a MALDI-TOF spectrum .....	132
Figure 33. Scheme of peptide quantification .....	134
Figure 34. Arg 0-2 and Arg 0-5 gels .....	135
Figure 35. Phospho-kinetics of Hsp8 .....	137
Figure 36. ESI MS analysis of peptide A (as described in Table 4) .....	138
Figure 37. ESI MS analysis of peptide B (as described in Table 4) .....	139
Figure 38. Phospho-kinetics of Enolase-1 .....	142
Figure 39. Hsc70 structure 1kax .....	146
Figure 40. Enolase-1 structure 1te6 .....	149
Figure 41. Arg 0-2 gel .....	152
Figure 42. Phospho-kinetics of Hsp4 .....	154
Figure 43. Phospho-kinetics of Hsp5 .....	156
Figure 44. Phospho-kinetics of CRMP-2 .....	159
Figure 45. Summary of MS analyses of RhoGDI-1, spot 221 .....	161
Figure 46. Phospho-kinetics of RhoGDI-1 .....	162
Figure 47. ESI MS analysis of peptide A (as described in Table 9) .....	165
Figure 48. ESI MS analysis of peptide B (as described in Table 9) .....	166
Figure 49. ESI MS analysis of peptide C (as described in Table 9) .....	167
Figure 50. ESI MS analysis of peptide D (as described in Table 9) .....	168
Figure 51. ESI MS analysis of peptide E (as described in Table 9) .....	169
Figure 52. RhoGDI-RhoA immunofluorescence .....	171
Figure 53. RhoGDI-Rac immunofluorescence .....	172
Figure 54. Phospho-kinetics of Rab1 and Rab6 .....	173
Figure 55. Phospho-kinetics of synthesis related proteins .....	175
Figure 56. Protein sequence alignment of Hsp8, Hsp5 and Hsp4 .....	179
Figure 57. RhoGDI-1 in complex with Cdc42 structure 1doa .....	186
Figure 58. Rab1 and Rab6 structures .....	188

## TABLES

Table 1. Proteins contained in the phosphoprotein fraction .....	87
Table 2. Distribution of proteins with acidic motifs and clusters .....	92
Table 3. List of proteins identified .....	107
Table 4. Peptides selected for the ESI MS/MS analysis of Hsp8 .....	137
Table 5. Proteins that act as chaperones/mediators of protein folding .....	153
Table 6. Proteins that act as metabolic enzymes .....	157
Table 7. Proteins involved in cell motility or structure .....	157
Table 8. Proteins that act as signalling molecules .....	158
Table 9. Summary of the peptides of interest and correspondent ions selected for ESI MS/MS analysis of RhoGDI-1 .....	164
Table 10. Proteins involved in protein synthesis.....	174

## ABBREVIATIONS

2D-PAGE	2 dimensional polyacrylamide gel electrophoresis
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BDNF	Brain-derived neurotrophic factor
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
cAMP	Cyclic adenosine-3',5'-monophosphate
CHAPS	(3-(3-Cholamidopropyl)-dimethylammonio-1-propanesulfonate
CDK	Cyclin dependent kinase
CID	Collision-induced dissociation
DAG	Diacyl glycerol
DMF	Dimethylformamide
DMSO	Dimethyl sulphoxide
DTT	Dithiothreitol
EDTA	Ethylene-diamine tetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
ESI	Electrospray ionisation
FBS	Foetal bovine serum
FHA	Forkhead-associated
FTICR	Fourier Transform Ion Cyclotron Resonance
GAPs	GTPase activating proteins
GEFs	Guanine exchange factors
GDI	Guanine dissociation inhibitor
GPCR	G-protein coupled receptor
Grb2	Growth factor receptor binding protein 2
GRP	Glucose related protein
GTP	Ganosine thriphosphate
HB-EGF	Heparin binding EGF-like growth factor

HBSS	Hank's balanced saline solution
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
HPLC	High pressure protein liquid chromatography
HSC70	Heat shock protein cognate 70
HSP70	Heat shock protein 70
IEF	Isoelectric focusing
IL-1	Interleukine 1
IMAC	Immobilised Metal Affinity Chromatography
IP	Immunoprecipitation
IP3	Inositol 1,4,5-trisphosphate
IPG	Immobilised pH gradient
JAK	Janus kinase
JNK	Jun N-Terminal kinase
LC	Liquid chromatography
LC/MS/MS	Liquid chromatography/tandem mass spectrometry
m/z	Mass-to-charge ratio
MALDI-TOF	Matrix-assisted laser desorption-ionisation Time-of -flight
MAPK	Mitogen activated protein kinase
MAPKAP K2	Mitogen-activated protein kinase-activated protein kinase 2
MBP1	Myc-binding protein 1
MEF-2	Myocyte enhancer factor 2
MEM	Minimal essential medium
MFP	Mass Fingerprinting
MNK-1	MAPK-1 interacting kinase 1
MKK	MAPK kinase kinase
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NBT	Nitro blue tetrazolium
NGF	Nerve growth factor
NEAA	Non-essential amino acids
NO	Nitrous oxide
NRK49F	Normal rat kidney 49F
NT-3	Neurotrophin 3

PBS	Phosphate buffered saline
PDI	Protein disulphide isomerase
PDK-1	Phosphoinositide-dependent kinase-1
PI3-K	Phospho-inositol 1,4,5-trisphosphate kinase
PIP2	Phosphatidyl-inositol 4,5-bisphosphate
PIP3	Phosphatidylinositol-3,4,5-trisphosphate
PTB	Phosphotyrosine binding
PTM	Post translational modification
PKA	Protein kinase A (cAMP-dependent kinase)
PKB	Protein kinase B
PKC	Protein kinase C
PLC	Phospholipase C
PLD	Phospholipase D
p90RSK	90 kDa ribosomal protein S6 kinase
RTK	Receptor tyrosine kinase
SAPK	Stress activated protein kinase
SDS	Sodium dodecyl sulphate
Shc	Src homology collagen protein
SH2	Src homology 2
SILAC	Stable isotope labelling with amino acid in cell culture
STAT	Signal transducers and activators of transcription
Sos	Son of sevenless
TAK1	Transforming growth factor $\beta$ - activated protein kinase 1
TBS	Tris buffered saline
TCA	Trichloroacetic acid
TEMED	Tetramethylethylenediamine
TGF	Transforming growth factor
TFA	Trifluoroacetate
TMPS	Triple membrane-passing signal
TPR	Tetratrico peptide repeat
TRAF-6	Tumour necrosis factor-receptor associated factor 6
WB	Western blotting

## **AKNOWLEDGEMENTS**

During the course of my PhD I have received generous support from many friends and colleagues. In particular I would like to express my gratitude to the people who have been closest to my efforts.

I would like to thank my supervisor Jasminka for her support in my studies and her confidence in me. Our many interactions have inspired my scientific thought and developed my broader skills.

I appreciate Julie's prompt feedback and her positive contribution to my work.

I am grateful to the members of my lab, past and present, for creating a pleasant and productive atmosphere. I thank Vukic for teaching me the first steps in the lab. In particular, I extend warmest thanks to Oliver, Jelena, Marie Anne and Kajsia for being open to scientific discussion and above all, for having become great, supportive friends. I thank Mark for lending a helping hand and always being attentive.

I also thank Tony for being a gracious and supportive presence. I thank the very nice people in his lab: they have all been fantastic.

Special thanks go to my main collaborators Marnie and Geraint, working with whom has broadened my thinking and filled me with energy and enthusiasm. It has been a genuine pleasure.

I thank Proteosys for partially funding my PhD, though the main merit goes to Jasminka for endeavouring to find alternative finances.

I also thank Dave and Cristina for entrusting me with a Post Doctorate position during the writing of my thesis and for never putting pressure on me.

I finally thank Toby for his advice, his patience and for choosing to put up with my moods.

Grazie mille a mamma e papà per non aver mai dubitato di me.



## POSTERS AND PUBLICATIONS

### Posters and talks

- “Enrichment Of Phosphoproteins For Proteomic Analysis Using Immobilized Fe(III)-Affinity Adsorption Chromatography”. Poster presented at the HUPO 4<sup>th</sup> Annual World Congress. 28<sup>th</sup> August -01<sup>st</sup> September 2005, Munich, Germany.
- “Quantitative Phosphoproteomics of EGF signalling”. Talk given at the ISMB Retreat. 27-28<sup>th</sup> June 2005, Cambridge, UK.

### Publications

- Guerrero I.C., Predic-Atkinson J., Kleiner O., Soskic V. and Godovac-Zimmermann J. 2005, “Enrichment of Phosphoproteins for Proteomic Analysis Using Immobilised Fe(III)-Affinity Adsorption Chromatography”. *Journal of Proteome Research*. In press.
- Guerrero I.C. and Kleiner O. 2005, “Application of Mass Spectrometry in Proteomics”. *Bioscience reports*. In press.
- Guerrero I.C.\*, Justin N.\*, Webb R.\*, Perez-Mansilla B., Tunwell R., Gamblin S., Godovac-Zimmermann J. and Thomas G.M.H. “Phospho-ADP-ribosylation factors (pARFs); a new family of GTP dependent four-way signalling switches”. (\* these authors contributed equally to this work). In preparation.

# **CHAPTER 1**

## **INTRODUCTION**

### **The role of proteomics in elucidating signalling networks**

The term PROTEOMICS refers to the analysis of all proteins in a living system, including the description of co- and post-translationally modified proteins and alternatively spliced variants. Together with protein expression analysis, proteomics includes the study of covalent and non-covalent protein associations and their spatial and temporal distributions within the cell. Ultimately, proteomics helps understand how biochemical and trafficking events are affected by changes in extracellular and intracellular conditions. Mass spectrometry (MS) is at the heart of virtually all proteomics experiments as it provides the key tools for the analysis of proteins. Developments in technology and methodology in the field of mass spectrometry and proteomics have been rapid over the last five years and are providing improved and novel strategies for global understanding of cellular function.

The history of proteomics dates back to the invention of two-dimensional gels in the 1970s, which provided the first feasible way of resolving and displaying hundreds or even thousands of proteins on a single gel. Identification of the spots separated on these gels increased enormously when biological mass spectrometry developed into a sufficiently sensitive and robust technique.

## **Mass spectrometry in proteomics**

### *Principles and instrumentation*

Mass spectrometers consist of three basic components: an ion source, a mass analyser, and an ion detector. MS measurements are carried out on ionised analytes in the gaseous phase, requiring a method to transfer molecules from solution or solid phase into this state. The two most commonly used techniques, which have been used in this study are: matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) (Fenn et al. 1989; Karas et al. 1988). Both MALDI and ESI are soft ionisation techniques in which ions are created with low internal energies and thus undergo little fragmentation. MALDI ionisation technique tolerates a reasonable amount of impurities in the sample to be analysed. Many different samples can be processed rapidly in an automated manner and they can be kept on the target for several days without compromising the quality of the analysis, allowing easy re-analysis when required.

ESI is based on spraying an electrically generated fine mist of ions into the inlet of a mass spectrometer at atmospheric pressure. This technique ionises molecules directly from solution, so it can easily be interfaced with liquid separation methods. Steady advances in the application of ESI to the analysis of peptides and proteins have been made, and the most notable improvements have come from reduction in the flow rate of the liquid used to create the electrospray, leading to more efficient creation of ions (Yates 2004).

After ionisation, the sample reaches the mass analyser, which separates ions by their mass-to-charge ( $m/z$ ) ratios. Ion motion in the mass analyser can be manipulated by electric or magnetic fields to direct ions to a detector, which registers the numbers of ions at each individual  $m/z$  value. In proteomics research, four basic kinds of mass analysers are currently being used: time-of-flight (TOF), ion trap, quadrupole, and Fourier transform ion cyclotron

resonance (FTICR) analysers. All four differ considerably in sensitivity, resolution, mass accuracy and the possibility to fragment peptide ions. The latter results in mass spectra with an especially high content of information (MS/MS spectra) (Aebersold et al. 2003).

The combination of ion source, mass analyser and detector is usually determined by the application. MALDI is usually coupled to TOF analysers, which separate ions according to their flight time down a field-free tube. Ions' time-of-flight is directly related to their  $m/z$  values and thus a mass spectrum can be acquired. The biggest drawback to TOF analysers is their inability to routinely perform MS/MS.

The ESI used for these experiments was coupled to ion traps (also called three-dimensional ion traps). Ions are first captured in the centre of the device for a certain time interval and are then scanned from the trap to the detector (Yates 2004). With this type of mass analyser it is possible to determine the mass of a given peptide as well as the sequence. Ions with specific  $m/z$  ratios can be selected in the 'trap' for fragmentation, induced by collision of the ion with an inert gas or a surface in a process called collision induced fragmentation (CID). This energy causes the peptide ion to fragment at different points, commonly at the peptide bond. The recorded ion product represents the tandem mass spectrum (MS/MS or  $MS^2$  spectrum)(Aebersold et al. 2001), which contain information on the amino acid sequence.

### *Protein identification*

The identification of proteins in proteomics is almost exclusively achieved by mass spectrometry (Aebersold et al. 2003). Current MS methods can identify any protein for which a primary structure is known. MS is highly sensitive: proteins present in femtomoles quantities can be identified on a routine basis and extension to low zeptomole amounts has been achieved (Shen et al. 2004; Trauger et al. 2004). Additionally, MS can in principle characterise

the exact covalent structure of a protein, allowing detection of splice variants and post-translational modifications, without prior knowledge of the types of modifications or their locations.

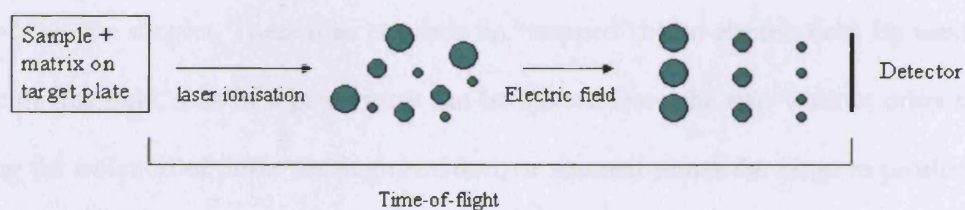
### *Mass fingerprinting (MFP)*

The most common MS strategies analyse peptides rather than full-length proteins. MALDI-TOF mass spectrometry is relatively simple and robust to operate, has good mass accuracy, high resolution and sensitivity. It is widely used in proteomics to identify proteins from simple mixtures by a process called peptide mass fingerprinting (Cottrell 1994).

In this approach, proteins of interest are digested with a sequence-specific enzyme such as trypsin. Tryptic peptides obtained are co-crystallised with an organic matrix on a metal target. A pulsed laser is used to excite the matrix, which causes rapid thermal heating of the molecules and eventually desorption of ions into the gas phase. An electrical field accelerates the ions towards the detector. Because of the usage of a pulsed laser, MALDI produces packets of ions rather than a continuous beam; it therefore needs to be coupled to a mass analyser that can measure either a complete mass spectrum without scanning a mass range, or trap all the ions for subsequent mass analysis. All the ions are given the same amount of kinetic energy so, because they differ in mass, they arrive at the detector at different times (a heavy ion will take longer to arrive than an ion with a smaller mass). If the machine is suitably calibrated with a range of known masses then the mass of the sample ions can be determined by the time they took to reach the detector. Trypsin is known to cleave proteins after a lysine or arginine residue. Therefore if the sequence of a protein is known, it is possible to predict what masses would be produced if that peptide were cleaved by trypsin. Several internet-based databases now exist in which the theoretical digests of known proteins are stored. By matching the spectrum of masses obtained by mass spectrometry, usually in the range of 800-3000 m/z,

with theoretical masses in the databases it is possible to identify the starting protein from a tryptic digest sample. This technique is known as “mass fingerprinting” (MFP) (Mann et al. 1993; Pappin et al. 1993). Searching the database provides a list of possible matches and a score for each indicating the probability that that protein is the correct match. The larger the number of peptides compared, and the more accurate the mass measurements made, the more likely it is that this method of protein identification will be successful. However, possible matches may have similar probabilities, and factors such as post-translational modification may interfere with database matching. If a clear match can not be found, or the location of a post-translational modification is to be pinpointed, the remaining sample can be sequenced using an ESI-Ion Trap mass spectrometer.

This form of protein identification ideally requires an essentially purified protein. The technique is therefore very often used in conjunction with two-dimensional gel electrophoresis (2D PAGE).



### Figure 1. MALDI-TOF Mass Spectrometry

The sample is ionised, then accelerated by the electric field. The heavier ions are slower to reach the detector than ions of lower mass. The mass of each ion can be calculated using the time it took to reach the detector. When looking at low-mass ions, a reflector is often used to increase the distance that the ions travel and thus improve the separation of similar masses.



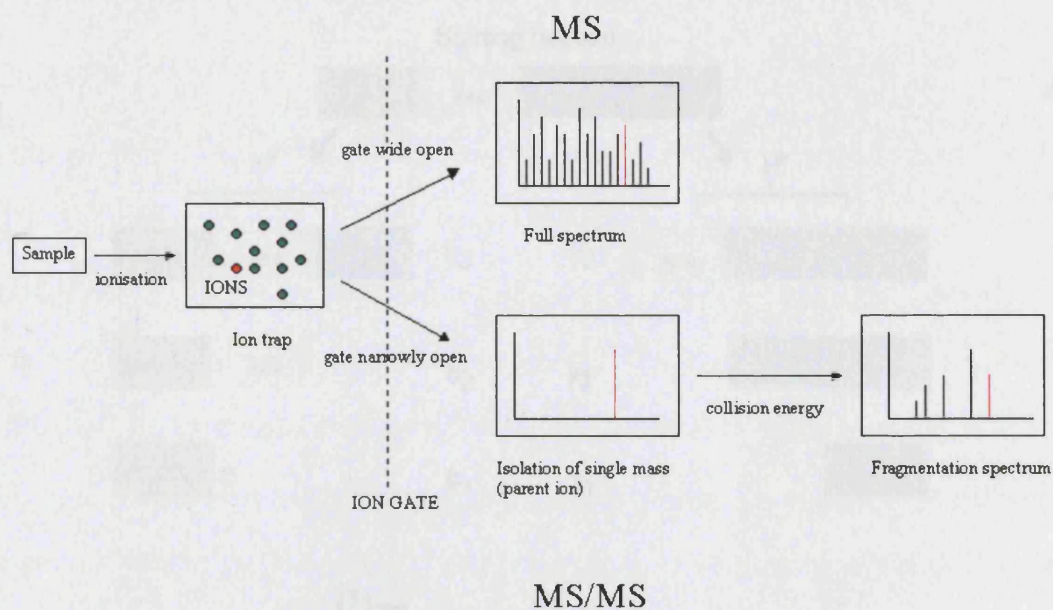
### *Protein sequencing: Tandem MS or MS/MS or MS<sup>2</sup>*

ESI ion trap and ESI quadrupole MS are currently the most popular setups for the simultaneous analysis of large numbers of peptides derived from enzymatic digests of complex protein mixtures. This is mainly due to their ability to interface chromatographic peptide fractionation methods online with the ESI based instruments and the possibility to obtain not only mass but sequence information as well. The other main advantage of ESI over MALDI is the production of multiply charged ions, which allows the measurement of high molecular weight peptides, often neglected in MALDI-TOF analysis.

Nano-electrospray is a variant of electrospray ionisation that, due to the slow introduction of sample into the machine, allows for the analysis of very small volumes of sample. This technique (in positive-ion mode) passes the sample, in 50% methanol, 0.5% acetic acid, through a very fine capillary while a high voltage is applied. This creates a spray of droplets, carrying the same net charge, which is passed through a heated capillary. While the mechanism of ionisation is not completely understood, it is thought that evaporation of the solvent causes the droplets to shrink and the charge density to increase. This brings the positive ions closer together, and eventually the force of repulsion between them is strong enough to cause the ions to leave the droplet. These ions can then be “trapped” by an electric field. By varying the voltage of this field, ions of a given mass can be ejected from the trap without other masses, allowing the isolation of peaks for fragmentation, or scanned across the range to produce a full spectrum of masses up to 2000  $m/z$  (Figure 2). The machine measures  $m/z$  (mass-to-charge ratio), as opposed to simply mass, so production of ions carrying multiple charges allows for the analysis of masses greater than 2000 Daltons.

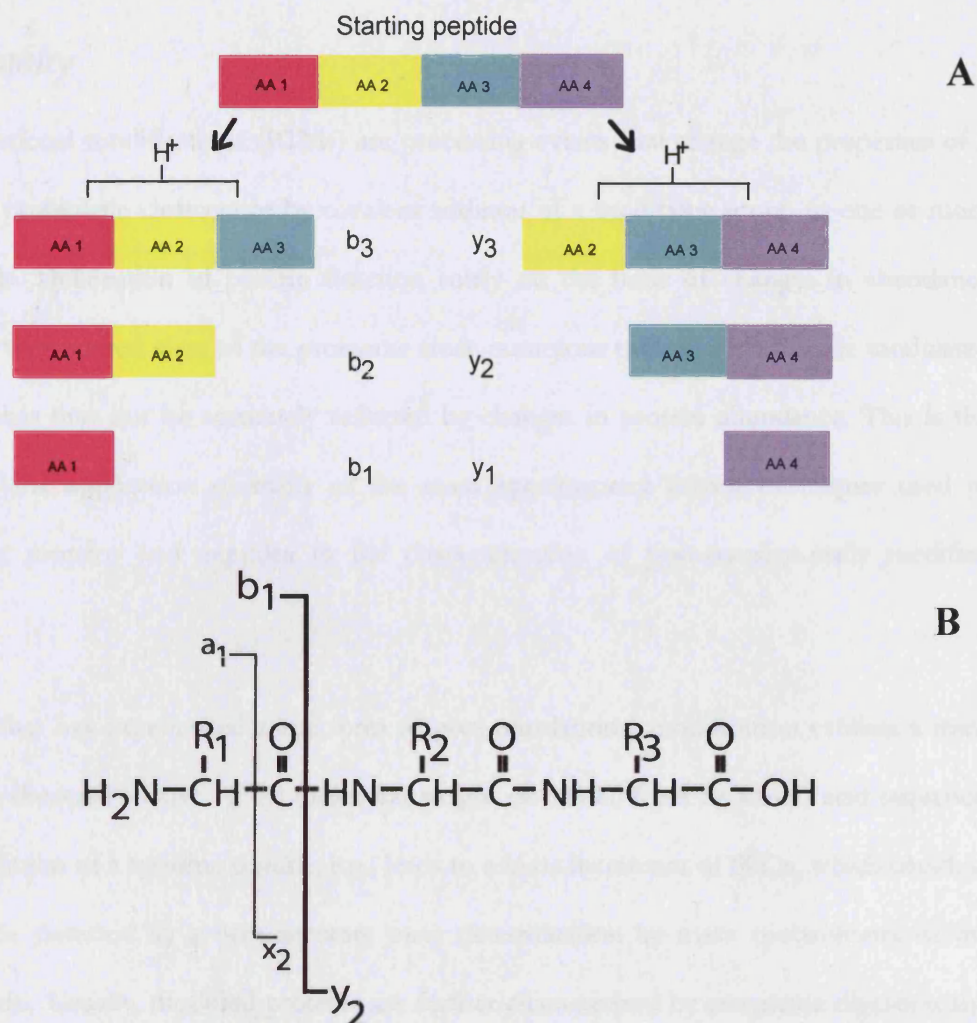
CID (Collision-Induced Dissociation) is the method by which a chosen peak can be sequenced. Ions of a particular mass are selected from the ion trap, and collided with atoms of

an inert gas. This can cause amino acids to be broken off the main chain of a peptide, producing a series of fragments (Figure 3). Adjusting the energy with which collisions take place varies the degree of fragmentation. Thus a single peptide can be examined at several collision energies, yielding a wide range of fragment sizes. A number of masses can be seen for each of these fragments. Groups such as  $-OH$  and  $-NH_2$  can be lost, as well as post-translational modifications like phosphate. The detector measures  $m/z$  so if a doubly charged ion is produced, a peak will appear at half the peptide mass, and so on. This gives rise to ions termed for example “ $a_2 - NH_3$ ” ( $a_2$  that has lost ammonia) or “ $a_2^{+2}$ ” ( $a_2$  carrying a double charge – this ion has an  $m/z$  of  $M+2H^+/2$ ) and permutations of these. The sequence information containing MS/MS spectra are compared against comprehensive protein sequence databases using one of several different algorithms (Eng et al. 1994; Mann et al. 1994; Perkins et al. 1999). This strategy is very similar to peptide mass fingerprinting and, in analogy, is called peptide fragmentation mass fingerprinting.



**Figure 2. ESI/Ion-Trap Mass Spectrometry**

After ionisation, the full range of ions can be displayed (MS), or an ion of interest (shown in red), can be isolated. Collision energy is then applied and the spectrum of the fragments produced is analysed (MS/MS).



### Figure 3. Peptide fragmentation by ESI MS/MS

(A) Collision-Induced Dissociation (CID). When collided with atoms of inert gas, the starting peptide is fragmented. When the spectrometer is used in "positive-ion mode", the sample is introduced in an acidic solution, so ionisation involves the addition of  $H^+$  to all fragments. "y" and "b" ions are ions generated from pieces of peptide sequence containing either the N-terminus (y) or the C-terminus (b). These ions are most useful for sequencing peptides. Other ions including one of the termini are known as "a", "x", "c" and "z" ions. (B) Ions can also be generated which include neither terminus. These are known as "internal fragments". An example would be an ion comprising AA2 and AA3 CID ion nomenclature. "y", "a" and "b" ions are the most useful for sequencing peptides. The nomenclature for these ions is related to where in the amino acid the dissociation occurs. How it corresponds to the peptide structure is illustrated above. "y" and "b" ions are reciprocals, as are "a" and "x" ions. All include either the C-terminus or the N-terminus.

## *Characterisation of post-translational modifications by mass spectrometry*

Post-translational modifications (PTMs) are processing events that change the properties of a protein by proteolytic cleavage or by covalent addition of a modifying group to one or more amino acids. Delineation of protein function solely on the basis of changes in abundance provides a very limited view of the proteome since numerous protein activities are modulated by PTMs that may not be accurately reflected by changes in protein abundance. This is the reason for the application of many of the mass spectrometry driven techniques used to characterise proteins and peptides to the characterisation of post-translationally modified proteins.

A protein that has experienced some form of post-translational modification exhibits a mass increase or decrease relative to the molecular weight calculated from its amino acid sequence. Phosphorylation of a tyrosine residue, e.g., leads to a mass increment of 80Da, which could, in principle, be detected by a very accurate mass determination by mass spectrometry of the intact protein. Usually, modified proteins are further characterised by enzymatic digestion and subsequent peptide mass mapping. In case the mass of the modified peptide is not sufficient to determine the nature of the modification and its location, the respective peptides are analysed by tandem mass spectrometry usually achieved by ESI MS.

These masses can be used to determine the sequence of a peptide. If a peptide is carrying a post-translational modification, its location can be pinpointed by observing the loss of that group from a series of backbone ions. Usually the mass of the ion with and without modification can be seen, but this is not always the case. The practical application of this technique is more clearly illustrated later in the results chapters by the sequencing of phosphopeptides of interest.



This procedure is very often the last experimental step in the characterisation of a modified protein or peptide. Problems may arise from the complexity of the post-translationally modified protein sample and the characteristics of the modified peptides (Stannard et al. 2003a; Stannard et al. 2003b). The size and the physicochemical properties of the peptides influence their ionisation and detection efficiency and consequently make them difficult to find in a high background of other ions. In addition, determination of the modified amino acid is sometimes not possible due to incomplete ion series in the tandem MS experiment.

Many of the currently used procedures to identify and characterise post-translationally modified proteins on a global scale employ some form of specific enrichment of the molecules to be identified prior to microcharacterisation by MS. This enrichment step can take place on the intact protein level as well as on the peptide level, as will be explained further in the next chapter.

## **Application of proteomics strategies in signalling studies**

### *The bottom-up and top-down approaches in proteomics*

Proteomics experiments most often invoke the simultaneous analysis of several thousands of protein from complex biological samples. The separation of peptides and proteins is therefore a key element in proteomics analysis providing a method to simplify complex mixtures and deliver molecules to the ionisation source. Most of the existing separation methods (including gel electrophoresis and liquid chromatography based methods) are currently in use and have been recently reviewed (Kleiner et al. 2005; Stannard et al. 2004; Wang et al. 2004; Yates 2004).

The final choice of the mass spectrometer and of the separation method to couple to it depends on the purpose of the study. This is the branching point of two main proteomics approaches, the bottom-up and the top-down approach.



In the bottom-up approach, complex protein mixtures are enzymatically digested into very complex peptide mixtures, which are then fractionated by multi-dimensional chromatography steps before they are subjected to tandem mass spectrometry. MS/MS spectra are recorded for as many peptides as possible, and the results used to search databases to identify the proteins in the original mixture. Such approaches are suitable for automation and high sample throughput can be achieved (Link 2002). These procedures typically identify a very limited number of peptides per protein, often just one, but still enough to identify the gene from which the protein was transcribed. They work well in the study of microorganisms where the assumption that one gene codes for only one protein species is often true. However when higher eukaryotes are investigated, processes including alternative splicing, RNA editing and post-translational modification can lead to several different protein species from a single gene. Peptide-based identification strategies enable the identification of the genes from which these proteins are derived but do not adequately identify the different functionally important protein isoforms (Godovac-Zimmermann et al. 2005; Link 2002).

In the top-down approach, intact proteins are displayed and different isoforms can be isolated before MS identification and characterisation. This is especially useful to unravel complex patterns of splice variations or post-translational modifications (Bogdanov et al. 2004; Kjeldsen et al. 2003; Reid et al. 2002; Sze et al. 2002). Difficulties in presenting intact proteins to the mass spectrometer, the complexity of the spectra and the lack of automation still restrict this application to time-consuming “one protein after the other” studies. A more common top-down approach is to fractionate and display intact proteins, but to analyse them on the peptide level. FT-ICR-MS broadens the applications of top-down proteomics, thanks to its capability to fragment intact protein *in situ* and to sequence them directly.

## ***Phosphoproteomics: the subproteome for signalling studies***

Cell signalling is often regulated by alteration of protein modifications. Proteins can have, amongst others, a modification in the shape of phosphate group, sulphate group, carbohydrates and lipids. Many post-translational modifications are reversible and can regulate the function of a protein. These modifications can not be predicted from gene sequence and are therefore subject to analysis of proteomics.

Protein phosphorylation is a major conduit of signal transduction within the cell. It has been intensively studied over the last decade since kinases regulate most of the signals originating at the cell surface by phosphorylation. Serine and threonine phosphorylation of proteins is more common, occurring in about 99% of cases, whereas tyrosine phosphorylation occurs in less than 1% of cases (Hunter 1998). This is also reflected in the genomic organisation of several organisms in which serine/threonine kinases far outnumber tyrosine kinases (Adams et al. 2000). In spite of the fact that a large number of kinases and phosphatases have been identified, very few physiological substrates have been described.

Cellular signalling appears to be more complex than previously speculated. The concept of one receptor for each signalling pathway is overtaken by many evidences proving that pathways from different receptors are interconnected and form a network of interacting proteins (Daub et al. 1997). In this network some proteins transfer the signals themselves by phosphorylation and dephosphorylation. Other proteins have no catalytic function but instead bring together components that are supposed to interact with each other. This theory of cell signalling where molecules are not dynamically moving around the cell, but rather form molecular assemblies through which signals are propagated is relatively new (Hur et al. 2002). In this way new groups of proteins, “scaffold” and “anchor” proteins, are introduced into the “club” of cell signalling members. Last year quite a few proteomics studies of protein interactions were

published. These experiments were designed to analyse protein complexes and they consist of affinity (bait) purification of the desired protein complex and identification of the bound proteins (Gavin et al. 2002; Ho et al. 2002).

Phosphorylation of signalling molecules can modulate their activities by changing their 3D structure, as has been noted in several kinases. Src homology 2 (SH2) and phosphotyrosine binding (PTB) domains specifically bind to phosphorylated tyrosine residues, whereas WW and Forkhead-associated (FHA) domains bind to phosphorylated serine and/or threonine residues (Pawson et al. 1997; Yaffe et al. 2001). The substrates known to date were mainly discovered using strategies that rely on *in vitro* phosphorylation activities of the kinase in question. Prediction of putative phosphorylation sites from so-called consensus sites generally yields a large number of false positives, making it necessary to confirm phosphorylation of residues *in vivo*.

In addition, there is emerging evidence of phosphorylation as a crucial determinant of functional partitioning of signalling components: the same molecule differently modified can play different roles in the cell (Stannard et al. 2003b).

### ***Phosphoproteomics tools***

Phosphoproteomics is a comprehensive analysis of protein phosphorylation that involves identification of phosphoproteins and phosphopeptides, localisation of the exact residues that are phosphorylated and quantification of phosphorylation. However, the analysis of phosphoproteins is relatively complex. Firstly, the stoichiometry of phosphorylation is generally relatively low as only a small fraction of the available intracellular pool of a protein is phosphorylated at any given time as a result of a stimulus. Secondly, the phosphorylated sites on proteins might vary, implying that any given phosphoprotein is heterogeneous (i.e. it exists

in several different phosphorylated forms). Thirdly, many of the signalling molecules are present at low abundance within cells and, in these cases; enrichment is a prerequisite before analysis. Fourthly, most analytical techniques used for studying protein phosphorylation have a limited dynamic range, which means that although major phosphorylation sites might be located easily, minor sites might be difficult to identify. Finally, phosphatases could dephosphorylate residues unless precautions are taken to inhibit their activity during preparation and purification steps of cell lysates.

Amongst the different methods used to target the phosphoproteome, anti-phosphotyrosine antibodies have widely used for immunoprecipitation of intact phosphoproteins in mass spectrometry driven proteomic studies of cell signalling (Blagoev et al. 2004; Soskic et al. 1999; Stannard et al. 2003a). However less specific, anti-phosphoserine and anti-phosphothreonine antibodies are used for the same purpose (Gronborg et al. 2002). A popular affinity-enrichment strategy for phosphopeptides in bottom- up proteomics approach is immobilised metal affinity chromatography (IMAC). Phosphate groups of phosphopeptides show a strong affinity towards immobilised  $\text{Fe}^{3+}$ - and  $\text{Ga}^{3+}$ -ions and this method is therefore used to enrich phosphopeptides from crude peptide mixtures prior to LC-MS/MS analysis in large scale phosphoproteomics experiments (Ficarro et al. 2002; Nuhse et al. 2003).

The application of this strategy into a top-down approach would require the isolation of the phosphoproteome using techniques like IMAC, but to date no publications describe an effective protocol to enrich phosphoproteins. In this study, focus has been given to the development of such procedure with the aim of separating the obtained phosphoproteins in a differential 2D display and identifying them. It is possible to sequence phosphopeptides in the mass spectrometer and unambiguously determine the site of phosphorylation (Papac et al.

1993; Roos et al. 1998), so some samples have been further analysed using ESI MS/MS, to obtain exact localisation of the phosphate in the protein sequence.

### *Quantitative mass spectrometry*

In addition to the initial identification and protein characterisation, a key parameter in proteomics analysis, is the ability to quantitate proteins of interest. Quantitation is a vital tool towards an understanding of transcriptional, translational and post-translational effects that affect protein production and function.

At this stage of knowledge in signalling, the quantification of every single phosphorylation site identified becomes the key and conclusive step. No such quantitative method applied to wide biological systems has been published to date.

There are several reasons why quantification of phosphorylation would be particularly important. A given protein might be in more than one signalling pathway (as is usually the case) with different stimuli inducing overlapping patterns of phosphorylation. That means a given site might not be phosphorylated at all, phosphorylated in a minority of molecules or, in an extreme case, on all the molecules of that protein. When a population of molecules from unsynchronised cells is analysed, this situation corresponds to detection of unphosphorylated, weakly phosphorylated or highly phosphorylated peptides containing the residue. Similarly, the ratio of phosphorylation of a protein on multiple residues might be crucial for its function.

In recent years quantitative proteomics by mass spectrometry has mainly focused on the differential quantitative determination of protein expression rather than absolute measurements, as many proteomic applications to drug target discovery or to track signalling events are concerned with relative rather than absolute abundances of proteins.

The techniques that have traditionally been used for quantification of phosphoproteins are based on  $^{32}\text{P}$  labelling of the cells, although MS-based techniques are rapidly evolving. Recently, stable isotope methods from small-molecule mass spectrometry have been introduced into MS-based proteomics and now allow relative changes to be determined. The principle of these methods is the incorporation of a stable isotope derivative in one of the two states to be compared. Stable isotope incorporation shifts the mass of the peptides by a predictable amount. The ratio of analyte between the two states can then be determined accurately by the measured peak ratio between the underivatised and the derivatised sample.

In mass spectrometry the amount of analyte in the sample does not correlate directly with the ion-current intensity of its mass spectrometric signal. Additional techniques have to be implemented to enable differential quantification of proteins with mass spectrometry. In proteomics almost all of these additional methods involve the labelling of peptides with stable isotopes by either biosynthetic or chemical methods. Peptides can then not only be identified but isotope labelling also allows the measurement of differential amounts of the same peptide (Lill 2003; Mason et al. 2003).

### *Metabolic isotopic labelling*

Metabolic labelling of proteins exploits the incorporation of isotopic labels during the process of cellular metabolism and protein synthesis. A simple and universal method for protein quantification by mass spectrometry is whole cell stable isotope labelling.

In the original approach, one experimental cell population is grown in an isotopically depleted medium enriched in  $^{15}\text{N}$ , and the other in standard  $^{14}\text{N}$ -rich medium. After extraction, the two protein samples are pooled and proteins/peptides are then analysed using classic proteomics methods (Oda et al. 1999; Pasa-Tolic et al. 1999; Washburn et al. 2002). The MS measurement

readily differentiates between peptides originating from the two pools because incorporation of a high abundance of  $^{15}\text{N}$  shifts the mass of any given peptide upwards, which leads to a pair of peaks from each peptide. The ratios between the intensities or areas of the lower and upper mass components of these pairs of peaks directly reflect the difference in the amount of a given protein in the two different cell pools. Accurate relative quantification is possible because the MS intensity response to a given peptide is independent of the isotopic composition of the nitrogen atoms (De Leenheer et al. 1992). An additional advantage of this approach is the availability of multiple peptide pairs per protein allowing a good estimation of the error associated with the quantification.

An increasingly popular method for protein quantification by MS is the labelling of proteins by incorporation of isotopically modified amino acids. This technique was recently named SILAC, for stable isotope labelling by amino acids in cell culture (Ong et al. 2002). In this approach, one population of cells is grown in medium containing the normal form of an essential amino acid while another population of cells is grown in medium supplemented with a stable isotope-labelled analogue. The two resulting protein samples are pooled and analysed essentially as described above. Different amino acids are suitable for SILAC, among others leucine (Leu-d0) and deuterated leucine (Leu-d3) (Ong et al. 2003), arginine ( $^{12}\text{C}_6$ -Arg) and fully substituted  $^{13}\text{C}$  labelled arginine ( $^{13}\text{C}_6$ -Arg) (Ong et al. 2002), tyrosine ( $^{12}\text{C}_9$ -Tyr) and fully substituted  $^{13}\text{C}$  labelled tyrosine ( $^{13}\text{C}_9$ -Tyr) (Ibarrola et al. 2004) have been used.

So far, metabolic isotope labelling has been used to compare two cellular states, such as diseased cells versus normal cells. Very recently the SILAC method has been extended to allow simultaneous quantification of proteins originating from three cellular states (Blagoev et al. 2004). Three different populations of cells were grown in media containing different versions of arginine,  $^{12}\text{C}_6$   $^{14}\text{N}_4$ -Arg,  $^{13}\text{C}_6$   $^{14}\text{N}_4$ -Arg and  $^{13}\text{C}_6$   $^{15}\text{N}_4$ -Arg. After protein extraction and pooling

of the samples, arginine-containing peptides generated by proteolytic cleavage occur as triplets in the mass spectra showing a mass separation by 6 or 10Da. Again, the intensity of each peak directly indicates the relative amount of the respective protein from the corresponding cell pool.

Until very recently, the use of metabolic labelling was limited to samples derived from cell culture. The appliance of this method has now been extended to quantitative  $^{15}\text{N}$  metabolic labelling of the multi-cellular organisms *Caenorhabditis elegans* and *Drosophila melanogaster* (Kjeldsen et al. 2003). Organisms can be labelled by feeding them on  $^{15}\text{N}$  -labelled *E.coli* and yeast, respectively, and the relative abundance of individual proteins was determined by mass spectrometry.

## **The Phospho-SILAC strategy**

In this introduction the main mass spectrometric tools available have been explained together with their applications in the investigation of signal pathways. Noticeably, most of the strategies used in proteomics to study signalling have been created to suit mainly bottom-up approaches. For instance, the study of phosphorylation has been mostly achieved by isolating phosphopeptides from mixtures of total proteins from one organism. As well, differential quantification has been largely developed to be applied to mixture of peptides rather than individual proteins.

Although the validity of results obtained using this approaches is not in dispute, one must be aware that these methods leave a whole level of information obscured. Analysing the peptide mixtures from the whole proteome is analysing peptides as protein surrogates. This allows identification of many of the proteins contained in the mixture and, occasionally, of one



possible phosphorylation site. No information is obtained on the number and the nature of proteins isoforms and their PTMs.

A more comprehensive, although not yet exhaustive, picture of the complicated signalling cascades can be only obtained by applying the newest technologies in a top-down approach. Indeed, in this thesis different methods were combined into a novel top-down tactic called the Phospho-SILAC strategy. First, a new phosphoprotein enrichment protocol was developed to isolate intact phosphoproteins rather than phosphopeptides. Second, an innovative application of the SILAC quantification method was developed. This focused on the quantification of changes of single phosphorylation at specific sites during EGF treatment of NRK49F cells. EGF signalling is described in the detail in the next sections.

Previous phospho-quantitative methods ( $^{32}\text{P}$  labelling, immunodetection) only profile the total level of phosphorylation of one given protein. In summary, the Phospho-SILAC method allows quantitative and simultaneous profiling of single phosphorylation sites within the same protein, which has not been achieved before. It also allows description of different isoforms of the same protein, and their difference in terms of phosphorylation sites and phosphorylation kinetics of each of these sites. When this strategy is applied to a large number of proteins, behaviour of some classes of proteins can be identified.

## **EGF signalling**

Signal transduction is essentially a mode of cell communication, and its mechanism is manifested through a chemical change, that directly or indirectly results in alterations in the activity or location of multiple proteins within a cell. In the past signal transduction pathways have been studied using a number of approaches including genetic, biochemical, and functional strategies. Most of the achievements have been based on the study of single

molecules to build up different networks. However, biochemical approaches to identify signalling molecules have in the past been hampered by the fact that large amounts of pure protein were generally required for identification by Edman degradation.

Proteomics makes use of classical methods but takes the study of cellular signalling a step further by employing 2D PAGE and/or various types of mass spectrometry, as described above. The major advantage of proteomics over classical methods is its ability to follow fluctuations of multiple proteins in parallel, rather than limiting itself to changes in individual proteins.

The study presented in this thesis is focused on the development and applications of the Phospho-SILAC proteomics approach to study EGF signal pathways. This work overcomes some of the limitations associated with classical biochemical approaches, such as focusing on one protein at a time. It also tries to surpass the restrictions of widespread proteomics approaches, which are often limited for the large amounts of data obtained and the lack of the necessary functional connections. EGF has been chosen as a signalling model because it is known to be very complex and therefore provided the challenge of contributing with quantitative and qualitative information to the established pathways and models.

## **EGF**

Epidermal growth factor (EGF) was discovered in the 1960s. A protein isolated from submaxillary glands of male mice was found to cause tooth eruption and premature eyelid opening in neonatal mice (Cohen 1962), and it was subsequently called epidermal growth factor based upon growth of epidermis in mouse tissues treated with it. For this work Stanley Cohen was awarded a Nobel Prize in Medicine together with Rita Levi-Montalcini in 1986.

EGF is a globular protein of 6.4 kDa consisting of 53 amino acids. It contains three intramolecular disulfide bonds essential for proper tertiary structure and therefore biological activity. EGF is found in varying concentrations in milk, saliva, urine (9-100 ng/mL), plasma (1 ng/mL) and most other body fluids. EGF proteins are evolutionary closely conserved; human EGF and murine EGF have 37 amino acids in common. Approximately 70 percent identity is found between human EGF and EGF isolated from other species, and the relative positions of the cysteine residues are conserved. Human EGF is synthesised as a very long prepro-protein of 1207 amino acids from which the factor (position 970-1023) is released by proteolytic cleavage (Ullrich et al. 1984). The EGF precursor is N-glycosylated and contains a hydrophobic domain which allows it to be anchored in the plasma membrane.

There is a whole family of EGF-like proteins, whose members are defined by high affinity binding to EGF receptor (EGFR), by structural homology and by production of mitogenic response in EGF-sensitive cells without being agonists of EGF. They contain so called EGF-like repeats, consisting of approximately 40 amino acids and characterised by a conserved arrangement of six cysteine residues (found in EGF itself) which are involved in the formation of three disulphide bonds. Betacellulin, heparin-binding EGF and transforming growth factor alpha (TGF $\alpha$ ) belong to this family.

EGF has been found to have many biological activities, the response to the stimulation depending on cell type and physiological circumstances. One of the most dramatic biological effects of EGF is the regulation of cell growth and differentiation. Other responses to EGF include: inhibition of the secretion of gastric acids (Gregory 1975), stimulation of vasoconstriction (Berk et al. 1985) and inhibition of differentiation (Kim et al. 1987).

EGF and EGFR have a critical role in multiple biological processes and detailed understanding of the signal transduction pathways and their crosstalks are of considerable interest. In

particular, the importance of this receptor is also due to the high levels of EGFR that have been observed in a variety of tumours and that have been correlated with disease progression, poor survival and poor response to therapy (Mendelsohn et al. 2000). A number of EGFR mutants have been observed in tumours where gene amplification has occurred (Kuan et al. 2001). One of the current EGFR-targeted strategies against tumours sees the use of monoclonal antibodies against the extracellular ligand-binding domain, in order to prevent ligand binding (Baselga 2002). A better understanding of the regulation of the downstream EGF signal pathway could suggest new anti cancer approaches.

### *EGF receptor*

EGF exerts its effects in the target cells by binding to the plasma membrane located EGFR. The latter belongs to the tyrosine kinases receptors (RTKs) family, a large class of proteins characterised by a single membrane-spanning domain and an intracellular tyrosine kinase catalytic domain. The extracellular chains are the least conserved, and vary considerably from one receptor to the other.

EGF receptors (EGFR) recognise six different structurally related ligands with high affinity and are expressed in almost all types of tissues at receptor densities of 20000-50000 copies/cell. The EGFR is synthesised from a 1210-residue polypeptide precursor; after cleavage of the N-terminal sequence, an 1186-residue protein is inserted into the cell membrane. The EGFR extracellular receptor domain (or ectodomain) has a length of 621 amino acids, including 11 glycosylated asparagine residues and 51 cysteine residues. This extracellular portion consists of four domains called L1, CR1, L2, and CR2 domains. Ligand binds between the L1 and L2 domains of the EGFR (Garrett et al. 2002; Ogiso et al. 2002). The CR1 and CR2 domains consist of a number of small modules, each held together by one or two disulfide bonds, and are responsible for dimer formation as a large loop protrudes from

the back of the CR1 domain and makes contact with the CR1 domain of the other receptor in the dimer.

The intracellular receptor domain has a length of 542 amino acids and encodes an intrinsic tyrosine-specific protein kinase activity (Ullrich et al. 1984). This kinase catalyses the transfer of the gamma-phosphate of ATP to a tyrosine residue of the receptor itself and to those of other intracellular proteins.

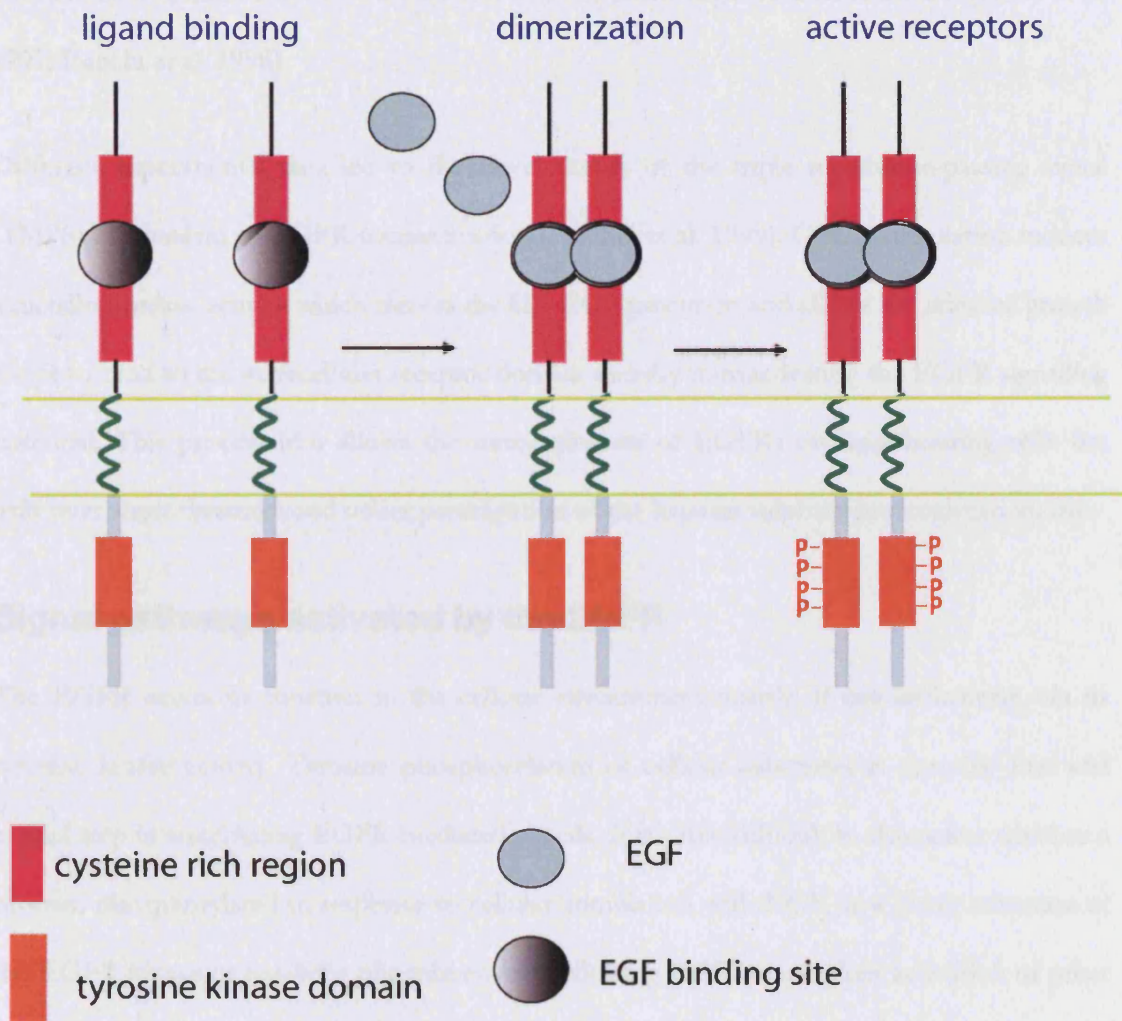
### ***EGFR oligomerization and activation***

The EGFR can form homodimers or interact with its three known homologues: ErbB2 (also called Neu or HER2), ErbB3 (HER3), and ErbB4 (HER4), in a ligand-dependent fashion to form heterodimers (Graus-Porta et al. 1997). Binding of EGF to the extracellular receptor domain activates the intracellular kinase domain of the EGFR. Each receptor molecule then phosphorylates the linked receptor on tyrosine, resulting in a interphosphorylation of the dimer in the C-terminal domain, leading to the activation of the dimer (Hubbard et al. 1998) (Figure 4). Recent studies have shown that the dimerisation is induced by the ligand, but mediated uniquely by the receptor-receptor contact (Ferguson et al. 2003). Differences in the C-terminal domains of the four different EGFR proteins result in changes to the repertoire of signalling molecules that interact with the heterodimers, thus leading to an expansion in the number of possible signalling pathways stimulated by a single ligand. The stoichiometry of the complex EGF-EGFR is 2:2 (Schlessinger 2000).

It has been suggested that the receptor monomers are in dynamic equilibrium with the receptor dimers, and only a limited population of receptor would exist dimeric. The ligand binding to the extracellular domain stabilises the formation of the active dimers. This model

would justify the enhanced activation of the protein kinase domain by receptor overexpression in the absence of ligand binding (Domagala et al. 2000).

The binding of ligand to the ectodomain releases the extracellular restraints on the formation of an active kinase dimer configuration. The monomeric EGFR has much reduced kinase activity compared to that of the dimerised receptor (Boni-Schnetzler et al. 1987; Sherrill 1997); it is assumed that in the absence of dimerisation, the kinase is in a relatively inactive conformation.



**Figure 4. Activation of EGF receptor**

Upon occupation of EGF, the EGF receptor forms a dimer and this induces a conformational change of the cytoplasmic domain that uncovers its latent tyrosine kinase activity. This phosphorylates the tyrosine residues of the bound receptor (interphosphorylation). The dimerised and phosphorylated receptors constitute the active molecule.

### *EGFR transactivation*

EGFR and its relative HER2 are rapidly tyrosine phosphorylated after stimulation with the GPCR agonists endothelin-1 (ET-1), lysophosphatidic acid (LPA) or thrombin. This transactivation of a receptor tyrosine kinase leads to extracellular signal regulated kinase (ERK) activation, induction of fos gene expression and DNA synthesis which are abrogated either by

selective EGFR inhibition or by expression of a dominant-negative EGFR mutant (Daub et al. 1997; Eguchi et al. 1998).

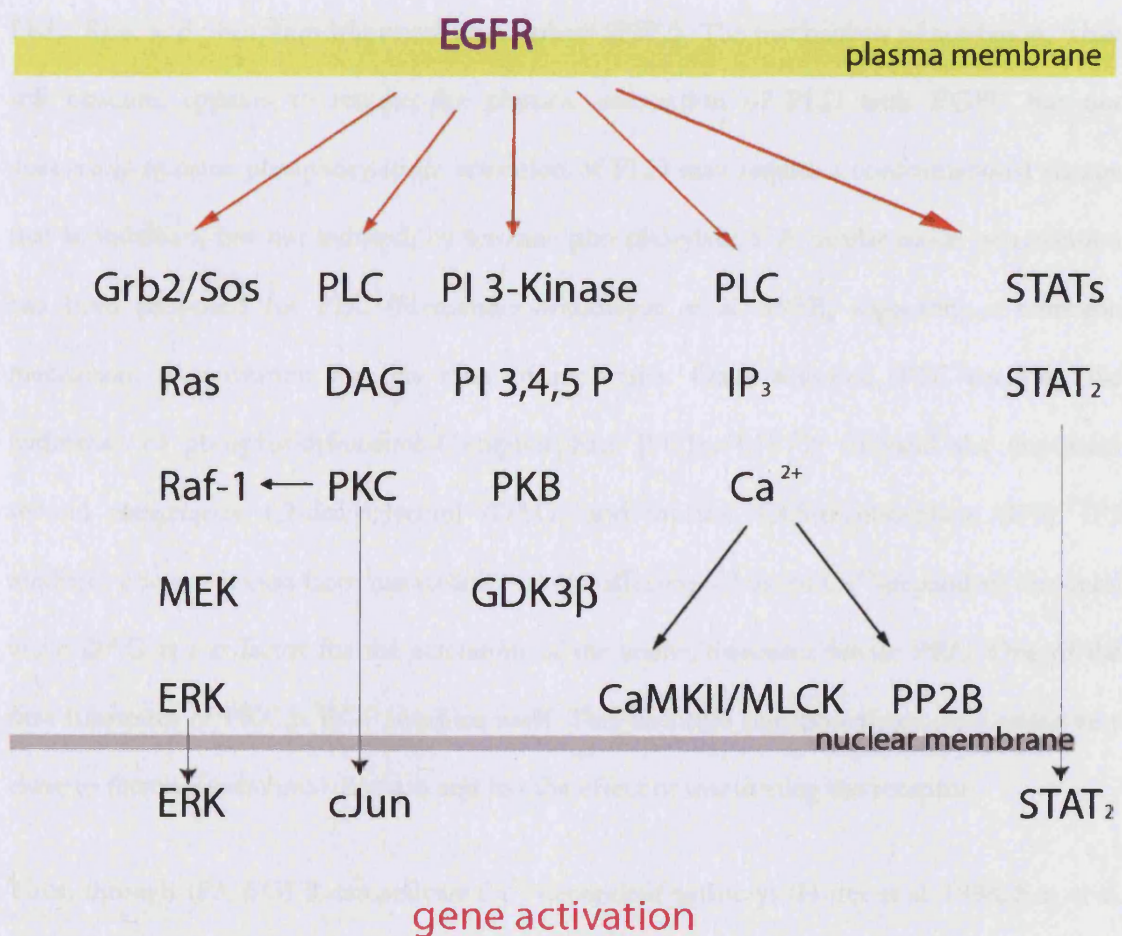
Different experimental data led to the novel theory of the triple membrane-passing signal (TMPS) mechanism of EGFR transactivation (Prenzel et al. 1999). GPCR stimulation induces a metalloprotease activity which cleaves the HB-EGF precursor and allows the released growth factor to bind to the extracellular receptor domain thereby transactivating the EGFR signalling potential. This process also allows the transactivation of EGFRs on neighbouring cells but only over short distances and under participation of the heparin sulphate proteoglycan matrix.

## **Signal pathways activated by the EGFR**

The EGFR exerts its function in the cellular environment mainly, if not exclusively, via its tyrosine kinase activity. Tyrosine phosphorylation of cellular substrates is thus the first and crucial step in transducing EGFR-mediated signals. It is often difficult to determine whether a protein, phosphorylated in response to cellular stimulation with EGF, is a direct substrate of the EGFR kinase or has been phosphorylated following EGFR-dependent activation of other cellular kinases. There are some well-known domains that mediate the physical association between EGFR and signalling proteins. Phosphorylation of the EGFR's C-terminus, be it autophosphorylation or trans-phosphorylation by other kinases such as Src and Jak-2 (Tice et al. 1999; Yamauchi et al. 1997), provides specific docking sites for the SH2, SH3 or PTB domains of intracellular signal transducers and adaptors, leading to their colocalization and to the assembly of signalling complexes. SH2 and PTB domain-mediated association of intracellular proteins with the EGFR are inducible and determined by the phosphorylation state of key tyrosine residues on the receptor. However, there are some proteins that are associated with the EGFR in its resting state, only to be activated or translocated to other cellular locations when the ligand binds (Xia et al. 2002).



Given the functional diversity of proteins that complex with, or are phosphorylated by, the EGFR, it is hardly surprising that EGF stimulation of a cell results in the simultaneous activation of multiple pathways often functionally interlinked. A scheme of some of the main pathways activated by EGFR is reported in Figure 5. More details on some of these pathways will be given in the next paragraphs.



**Figure 5. Branching of the signal transduction pathways.**

Following activation of EGFR, several signal transduction pathways can be activated. Five of these are indicated, and they are further explained in the text.

### *Phospholipids metabolism: PLD, PLC, and PI3-K*

EGF stimulation of a cell has marked effects on its phospholipid metabolism. Of the enzymes involved in these pathways, at least three can be activated directly by the EGFR, phospholipase C-gamma (PLC), phosphatidylinositol-3-kinase (PI3-K), and phospholipase D (PLD), while others are regulated indirectly by EGF-mediated activation of other pathways.

Until recently the stimulation was thought to be indirect and mediated by cofactors such as PKC, Rho, and phosphatidylinositol biphosphate (PIP<sub>2</sub>). The mechanism of activation, while still obscure, appears to require the physical association of PLD with EGFR but not necessarily tyrosine phosphorylation: activation of PLD may require a conformational change that is stabilised, but not induced, by tyrosine phosphorylation. A similar mode of activation has been proposed for PLC (Hernandez-Sotomayor et al. 1993), suggesting a common mechanism of activation for this class of molecules. Once activated, PLC catalyzes the hydrolysis of phosphatidylinositol-4,5-biphosphate [PtdIns(4,5)-P<sub>2</sub>] to yield the important second messengers 1,2-diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP<sub>3</sub>). IP<sub>3</sub> mediates calcium release from intracellular stores, affecting a host of Ca<sup>2+</sup>-dependent enzymes, while DAG is a cofactor for the activation of the serine/threonine kinase PKC. One of the first substrates of PKC is EGF receptor itself. This becomes phosphorylated on a serine very close to the transmembrane domain and has the effect of inactivating the receptor.

Thus, through IP<sub>3</sub>, EGFR can activate Ca<sup>2+</sup>-dependent pathways (Hofer et al. 1998; Sun et al. 1998), and through PKC multiple signalling components, including the MAPK and JNK pathways (Marais et al. 1998) and possibly the Na<sup>+</sup>/H<sup>+</sup> exchanger (Liaw et al. 1998).

Phosphoinositide-3-kinases are major players in cellular functions, where they contribute to a variety of cellular processes including proliferation, survival, adhesion, and migration (Cantley 2002).

PI3-K generates phosphatidylinositol-3,4,5-trisphosphate (PIP3). One of the best characterised targets of this second messenger is the Ser/Thr kinase PKB (Jones et al. 1991), which binds to the lipid and is translocated to the plasma membrane where it is phosphorylated and activated by phosphoinositide-dependent kinase-1 (PDK-1). PKB is a major mediator of PI3-K action in survival and proliferation, and may well be the major mediator of the antiapoptotic effects of EGFR activation.

### ***MAPK (Mitogen activated protein kinase) pathway***

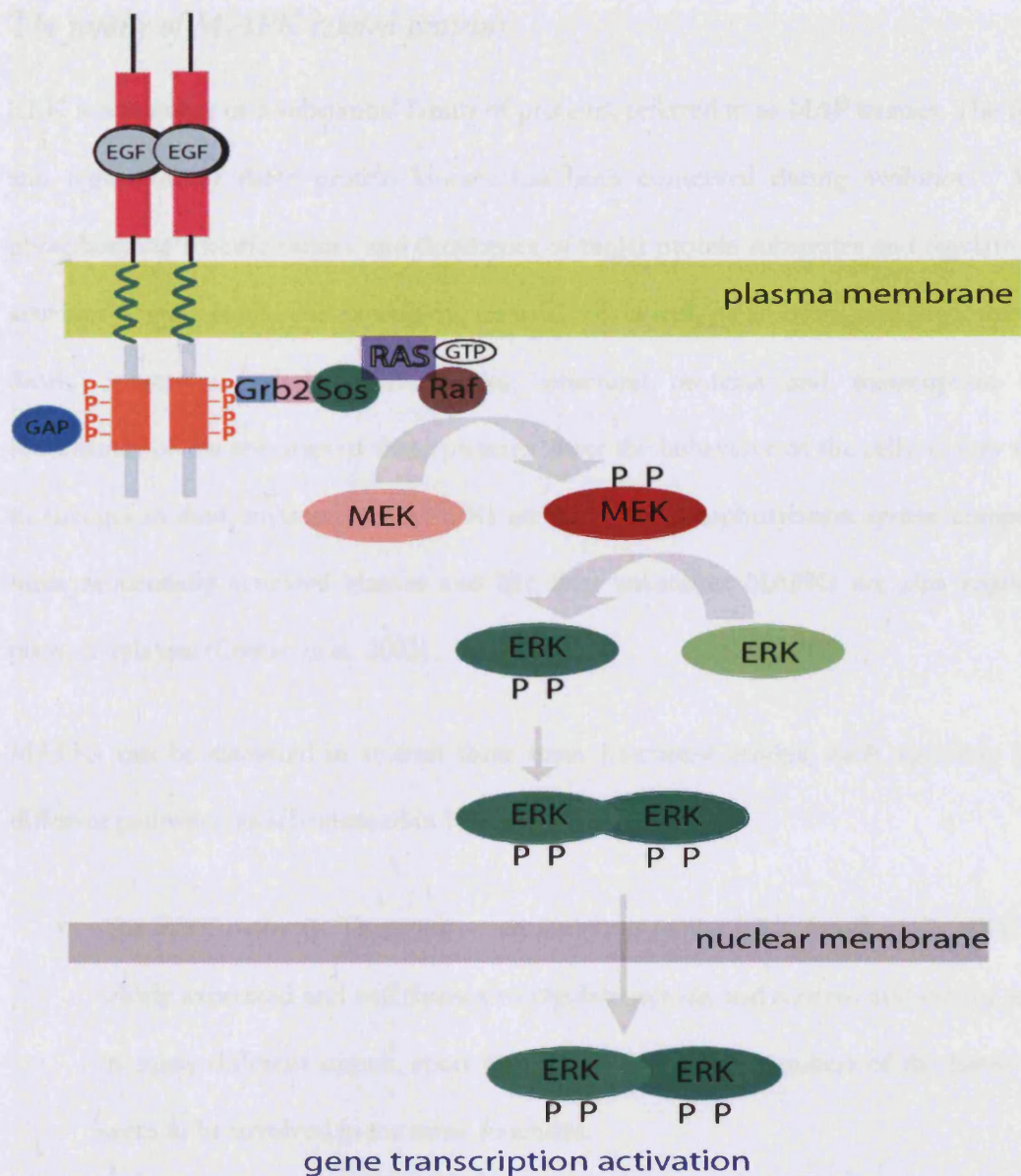
#### ***Grb2/Ras pathway***

The cascade of biochemical events that leads from the EGFR to the activation of the proto-oncogene Ras and, eventually, the serine/threonine kinase MAPK has been analysed extensively. The key player in EGF-dependent Ras activation is the adaptor protein Grb2 (Growth factor Receptor Binding protein) (Lowenstein et al. 1992), which is constitutively bound to the Ras exchange factor Sos (son of sevenless) and is normally localised to the cytosol. Following activation of the EGFR kinase and autophosphorylation, the SH2 domain of Grb2 can bind to the EGFR. It must be noted that Grb2 can associate with the receptor either directly via Tyr1068 and Tyr1086 (Sorokin et al. 1994), or indirectly, by binding to EGFR-associated, tyrosine phosphorylated Shc (Sasaoka et al. 1994). It has been suggested that association of Shc to EGFR via its PTB domain, leading to its tyrosine phosphorylation and to the recruitment of Grb2, is the main step in EGF-dependent induction of the Ras/MAPK pathway (Sakaguchi et al. 1998). Activated Ras in turn activates the

serine/threonine kinase Raf-1 (Hallberg et al. 1994). Raf-1 activation, through a series of intermediate kinases, such as MEK (MAPK-ERK Kinase), leads to the phosphorylation, activation, and nuclear translocation of Erk-1 and Erk-2 (Extracellular signal Regulated protein Kinase), which catalyse the phosphorylation of nuclear transcription factors (Johnson et al. 1994). Activation of the MAP kinases also provides a negative feedback loop for this pathway since the Grb2-Sos complex is dissociated following MAPK phosphorylation of Sos (Langlois et al. 1995).

Both Grb2 and Shc play important roles in the activation of other EGFR-dependent pathways. They are a beautiful example of “modular” construction. Grb2 contains two SH2 domains and one SH3 domain, which enables it to interact with tyrosine-phosphorylated motifs as well as with proline-rich regions of other proteins (Meisner et al. 1995). Shc can associate with specific tyrosine-phosphorylated sequences via its SH2 and PTB domain, and, being itself phosphorylated on tyrosine by activated receptors and cytosolic tyrosine kinases, serves in turn as a binding partner for SH2-containing proteins. The presence of such common domains sets the primary structure analysis of novel proteins as a valuable tool to attribute them an involvement in the signalling network.





**Figure 6. Regulation of Ras/MAPK pathway by EGFR**

The adaptor protein Grb2, in association with the guanine exchange factor Sos, attached to the tyrosine phosphorylated receptor through its SH2 domain. This brings the Grb2/Sos complex into the vicinity of the membrane where it catalyses the guanine exchange GTP/GDP exchange on Ras. Activated Ras associate with serine/threonine kinase Raf-1. Its localisation at the membrane results in activation and phosphorylation of MEK, which phosphorylates ERK on both tyrosine and threonine residues. Dimerisation exposes a signal peptide that allows translocation to the nucleus.

### *The family of MAPK related proteins*

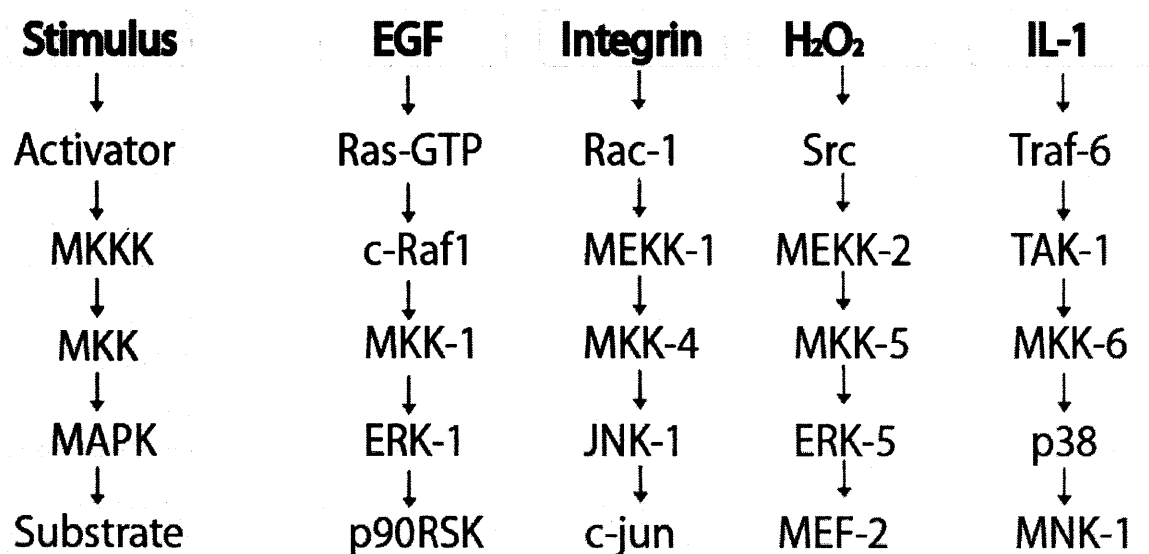
ERK is a member of a substantial family of proteins, referred to as MAP kinases. The function and regulation of these protein kinases has been conserved during evolution. MAPKs phosphorylate specific serines and threonines of target protein substrates and regulate cellular activities ranging from gene expression, mitosis, movement, metabolism, and programmed cell death. Substrates include other kinases, structural proteins and transcription factors. Modulation of the activities of these proteins alters the behaviour of the cells as they respond to changes in their environment. MAPKs are part of a phosphorylation system composed of three sequentially activated kinases and like their substrates MAPKs are also regulated by phosphorylation (Cowan et al. 2003).

MAPKs can be classified in at least three main functional groups, each operating in three different pathways, as schematised in Figure 7:

- The ERK pathway. There are seven members of the ERK family (1-7). ERK1/2 are widely expressed and well known to regulate mitosis and meiosis and can be activated by many different stimuli, apart from EGF. The other members of the family do not seem to be involved in the same functions.
- The JNK/SAPK pathway. SAPK stands for Stress Activated Proteins Kinase. The JNK 1-3 (Jun N-Terminal kinase) are mainly involved in response to stress and inflammatory cytokines. They are known to bind and phosphorylate c-Jun, a DNA binding protein, and to enhance its transcriptional activity that indirectly regulates the expression of many cytokines.

- The p38 pathway. P38MAPK ( $\alpha, \beta, \gamma, \delta$ ) regulate the expression of many cytokines, and is involved in immune response. P38 $\alpha$  is the first MAPK that breaks the paradigm according to which the MAPKs are activated by MKK (Map Kinase Kinase). Recent work show that P38 $\alpha$  is activated by TAB1, a scaffolding protein with no enzymatic activity (Ohkusu-Tsukada et al. 2004; Tanno et al. 2003).

The role of scaffolding protein in signal transduction is becoming clearer and clearer, as I will describe in the following section: Chaperones involvement in EGF signalling.



**Figure 7. MAPK phosphorylation systems**

The modules shown in pink are representative of pathway connections for each MAPK family, whilst the module in light blue represents the general pathway. P90RSK: 90 kDa ribosomal protein S6 kinase. Src: an oncogenic tyrosine kinase. MEF-2: myocyte enhancer factor. IL-1: interleukin 1. TRAF-6: tumour necrosis factor-receptor associated factor 6. TAK1: transforming growth factor  $\beta$ - activated protein kinase 1. MNK-1: MAPK-1 interacting kinase 1.

## ***The JAK and STAT pathways***

EGF-mediated regulation of transcription can be effected, at least in part, by STAT-1, 3, and 5. STAT proteins are inactive transcription factors, which are activated and translocated to the nucleus upon specific receptor stimulation.

Classically, STATs are recruited to the intracellular domain of the cytokine receptors through specific binding between STAT SH2 domains and receptor phosphotyrosine residues and are activated by JAK kinase (Kloth et al. 2003). The mode of activation by EGFR appears to be significantly different from that used by cytokine receptors. Firstly, the ligand-dependent phosphorylation of STATs by EGFR does not require JAK kinases (David et al. 1996); secondly, STATs do not bind to the C-terminal phosphotyrosines of the EGFR; indeed it appears that STATs are constitutively associated with the EGFR in its inactive state (Xia et al. 2002). However, as in JAK kinase signalling, activation of STAT transcriptional activity is strictly dependent upon the EGFR tyrosine kinase activity (David et al. 1996). More recent reports have also implicated the Src kinase in EGF-dependent STAT activation (Kloth et al. 2003; Olayioye et al. 1999).

## ***GTPases in signalling***

GTP-binding proteins (G-proteins) are important molecular switches that regulate a multitude of biological processes induced by extracellular signals by changing their status from inactive GDP-bound to active GTP-bound forms. Control of this transition is assisted by several different classes of proteins. The regulators for the Rab and Rho/Rac subfamilies include the guanidine nucleotide dissociation inhibitors (GDIs), which inhibit the dissociation of the nucleotide bound to these G-proteins. GDIs are known to extract the membrane-associated GDP-bound proteins from the membrane by binding to its prenylated C-terminus, keeping the Rho protein inactive in the cytosol by protecting the hydrophobic tail from aqueous solvent



(Isomura et al. 1991). Three RhoGDIs have been reported to act on proteins in the Rho/Rac family, RhoGDI-1, which is ubiquitously expressed and binds to a broad range of Rho/Rac proteins, RhoGDI-2, which is specifically expressed in haematopoietic tissues and associates with still uncharacterised Rho proteins, and RhoGDI-3, which is expressed in brain, lung and testis, and preferentially binds RhoB (Olofsson 1999).

Rho proteins and their regulatory proteins are involved in EGF signalling. In human lens epithelial cells, stimulation with EGF for 10-60 minutes leads to Rho-GTPase and Rac-GTPase activation of 45% and 65% respectively (Maddala et al. 2003). In gastrointestinal epithelial cells, microinjected RhoGDI can inhibit EGF dependent cell migration (Santos et al. 1997). The full molecular mechanisms for regulation of RhoGDI-1 are not yet clear, although several studies suggest that the binding affinity of different RhoGDIs for Rho-GTPase is regulated by selective phosphorylation of the GDIs (Bourmeyster et al. 1996; Gorvel et al. 1998). More recent studies suggest that RhoGDI-1 phosphorylation and Rho activation require PKC $\alpha$  activation (Mehta et al. 2001).

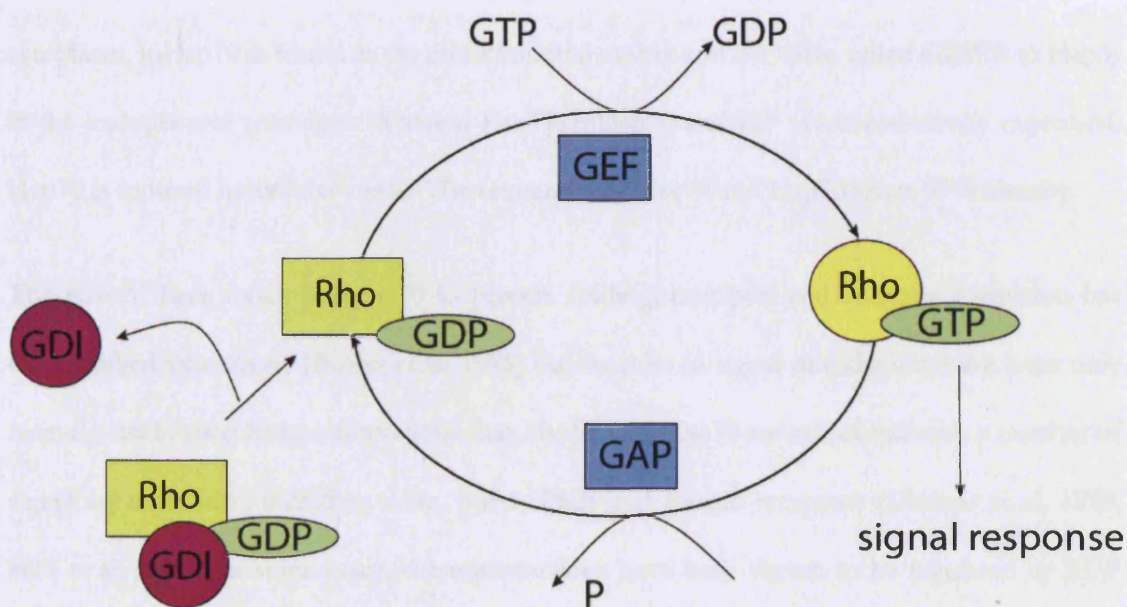
Rabs act as molecular switches as well, cycling between “inactive” GDP-bound and “active” GTP-bound states (Seabra et al. 2004) (Figure 8). Importantly, these changes in activity are coupled to reversible association with their target membranes. Whilst the GDP/GTP switch is a well-recognised molecular mechanism for controlling the activity of small GTPases, the membrane association/dissociation cycle, which could be described as a membrane in/out switch, is just as critical for the proper functioning of Rab proteins; to be in a truly activated state, a protein must be both GTP-bound and membrane-associated. Rab proteins associate to the membrane by virtue of the double prenylation on their C termini. This cycle of events allows both spatial and temporal control of Rab activity and is coordinated by several factors.

Geranyl-geranylated Rabs are delivered to their target membranes by a Rab escort protein (REP) or a Rab GDP dissociation inhibitor (RabGDI).

Rab proteins form the largest branch of the Ras superfamily of GTPases, with 63 different mammalian Rab proteins, many of which are restricted to specialised cell types (Seabra et al. 2004). They are localised to the cytoplasmic face of organelles and vesicles involved in the biosynthetic/secretory and endocytic pathways in eukaryotic cells. It is now well established that Rab proteins play an essential role in the processes that underlie the targeting and fusion of transport vesicles with their appropriate acceptor membranes. Rab1 controls transport events through early Golgi compartments and recent work suggests a role for Rab6 in intra-Golgi transport (Martinez et al. 1998). Trafficking of epidermal growth factor receptor (EGFR) involves small GTPases of the Rab families and Rab5 in particular has been shown to have a clear link with internalisation of the EGFR (Lanzetti et al. 2000). Subsequent studies have also indicated that Rab5 and the Raf/Erk signal transduction pathway play essential and selective roles in EGF-induced cell proliferation, and highlight a new function for Rab5 in EGF signalling (Barbieri et al. 2004).

Other aspects of EGFR trafficking are better established. Unoccupied EGFRs are mostly localised in caveolae, which represent a minimal part of the plasma membrane. These microdomains are characterised by a high concentration of gangliosides, sphingomyelin, cholesterol and caveolin (Smart et al. 1999). Following the addition of EGF and receptor activation, EGFR-EGF complexes move out of caveolae into clathrin-coated pits. These microdomains are constituted by clathrin triskelions, the assembly of which seems to be stimulated by phosphorylation of the heavy chain of the protein (Connolly et al. 1984). There is some evidence that Src might be involved in clathrin phosphorylation (Ware et al. 1997). Coated intracellular vesicles are then formed and subsequently the clathrin coat is removed to

produce early endosomes. This compartment then matures to form multivesicular bodies, where together with the early endosomes, the EGFRs can be either sent to the lysosome for degradation or they are dismantled into monomers and recycled to the membrane (Sorkina et al. 1999).



**Figure 8. GTPases cycle**

Rho GTPases cycle between an inactive GDP-bound form and an active GTP-bound form. The cycle is tightly regulated mainly by guanine exchange factors (GEFs), GTPase activating proteins (GAPs) and guanine dissociation inhibitors (GDIs). In their active form, Rho GTPases can bind to signalling molecules such as kinases and scaffold proteins and trigger the signal response.

### *Chaperones involvement in EGF signalling*

Chaperone molecules bind proteins in non-native states, preventing their aggregation and assisting them to reach a functional conformation in stressed cells. They also play important roles in cells under normal conditions, in folding newly synthesised and translocated proteins, disassembling oligomeric protein structures, facilitating proteolytic degradation and controlling the biological activity of folded regulatory proteins (Townsend et al. 2003).

The 70-kDa heat shock proteins constitute one of the best characterised families of molecular chaperones (Mayer et al. 2001). Although 70-kDa heat shock proteins have long been known to be involved in protein folding, they have been recently recognised as having important roles in cellular nucleo-cytoplasmic transport and in signal transduction. There are four major mammalian 70-kDa heat shock proteins, Hsc70 and Hsp70 which reside in the nucleus and cytoplasm, mHsp70 is found in the mitochondrial matrix and BiP (also called GRP78 or Hsp5) in the endoplasmic reticulum. Whereas Hsc70, mHsp70 and BiP are constitutively expressed, Hsp70 is induced by cellular stress. The sequences of Hsc70 and Hsp70 share 99% identity.

The role of heat shock protein 70 in protein folding, transport and complex formation has been studied extensively (Bukau et al. 1998) but its roles in signal transduction have been only recently established from observations that Hsp90 and Hsp70 are associated with a number of signalling molecules, including v-Src, Raf-1, PKB and steroid receptors (Dittmar et al. 1998; Sato et al. 2000). In some cases, these interactions have been shown to be regulated by EGF (see below for more details).

Hsp70/Hsc70 is composed of three main domains: an ATPase domain at the N-terminal, a peptide binding domain in the centre, and a tetratricopeptide repeat (TPR) binding motif at its C-terminal (Brive et al. 2002). Interactions with co-chaperones happen via Hsp70 N-terminal, while it binds to substrates via its C-terminal. Binding of exposed stretches of hydrophobic residues in unfolded proteins is regulated by ATP-hydrolysis induced conformational changes in the ATPase domain: release of substrate requires the binding of ATP to Hsp70, after which the substrate either is folded correctly or enters a new cycle of binding and release. Co-chaperones can influence the cycle of binding and release of Hsp70 with its substrate, by stabilising the ADP state and therefore increasing its chaperone activity, like Hip, (Hohfeld et al. 1995) or by accelerating the nucleotide exchange and consequent release of non mature

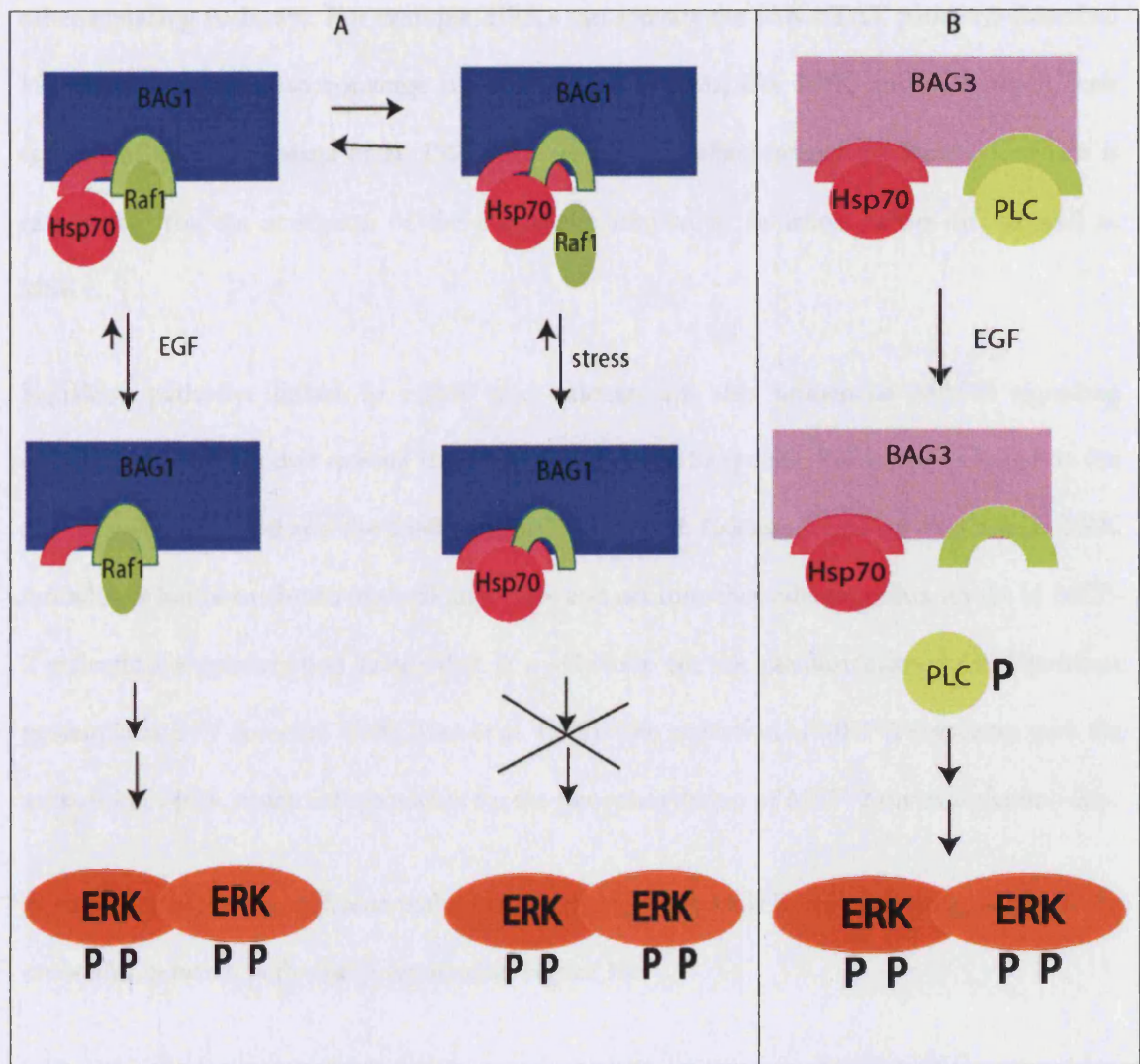
substrates, like Bag1 (Bimston et al. 1998; Brive et al. 2002). The ATPase domain is bi-lobed with a B-terminal lobe (residues 1-189) and a C-terminal lobe (residues 190-373) each consisting of separately folded units and the nucleotide binding crevice is between the lobes, and the nucleotide exchange happens in this crevice. It has been shown that the C-terminal lobe of the ATPase domain is enough to guarantee the binding with Bag domain (Briknarova et al. 2001; Brive et al. 2002; Song et al. 2001). In addition, Hip protein seems to bind regions 240-260 and 280-290 (Velten et al. 2000), in the C-terminal region of the ATPase domain of Hsp70.

Both Hip and Bag1 bind on the same site of the ATPase domain, and compete to influence Hsp70 activity (Nollen et al. 2001). Bag-1 and Hip also interact with other proteins in the cell. Bag1, for example, interacts with and influences the function of many key components of cell death and signal transduction pathways, including the anti-apoptotic protein Bcl-2 and the growth regulator Raf-1 (Takayama et al. 2001; Wang et al. 1996). It has been proposed that the regulation of Bag1 function by Hsp70 serves as a checkpoint to regulate growth and death when the levels of Hsp70 in the cell rise in response to stress (Song et al. 2001). Bag1, a co-chaperone for Hsp70, coordinates signals for cell growth in response to cell stress, by downregulating the activity of Raf-1 kinase. Raf-1 and Hsp70 compete for binding to Bag1, such that Bag1 binds to and activates Raf-1, subsequently activating the downstream ERK1/2. When levels of Hsp70 are elevated after heat shock, or in cells conditionally overexpressing Hsp70, Bag1–Raf-1 is displaced by Bag1–Hsp70, and DNA synthesis is arrested. Bag1 mutants which can not bind Hsp70 constitutively activate Raf-1/ERK kinases but are unaffected by Hsp70; consequently neither Bag1–Raf-1 nor DNA synthesis is negatively affected during heat shock. Likewise, Hsp70 mutants which retain chaperone activity but do not bind to Bag1, fail to repress Bag1 activation of Raf-1/ERK kinase. Bag1 functions in the heat-shock response to coordinate cell growth signals and mitogenesis, and that Hsp70

functions as a sensor in stress signalling. It must be added though, that this might only be one of the many mechanisms of regulation, because the levels of Hip and Bag1 are only 1% of that of Hsp70 (Figure 9, A).

Connections between Hsp70 family of proteins and PLC pathways have also been reported. BAG3 forms an EGF-regulated ternary complex with Hsc70 and PLC. Phosphorylated BAG3 binds to Hsc70 in intact cells and BAG3. PLC binds to BAG3 in a separate region in unstimulated cells, generating a ternary complex. Upon stimulation with EGF, PLC dissociates from the complex, whilst Hsc70 remains attached to BAG3. It is thought that Hsc70 modifies the 3D structure of PLC (kept in the vicinity by Bag3) and then releases the activated PLC upon BAG3 phosphorylation on tyrosine (Doong et al. 2000). This phosphorylation requires PKC, since it is abolished in the presence of PKC inhibitors (Figure 9, B).





**Figure 9. Scheme of Hsc70 involvement in Raf-1/ERK pathway and in PLC pathway**  
 (A) Hsp70 and Raf-1 have overlapping binding sites on BAG1. In case of stress, high abundant Hsp70 displaces Raf-1. Upon EGF stimulation, Raf-1 preferably binds to BAG1 and becomes activated (in a Ras-independent pathway). (B) Hsp70 and PLC have distinct binding sites on BAG3. Within 1 minute of EGF stimulation, PLC is activated and dissociates, whilst Hsp70 remains associated to BAG3.

### *Signalling crosstalk*

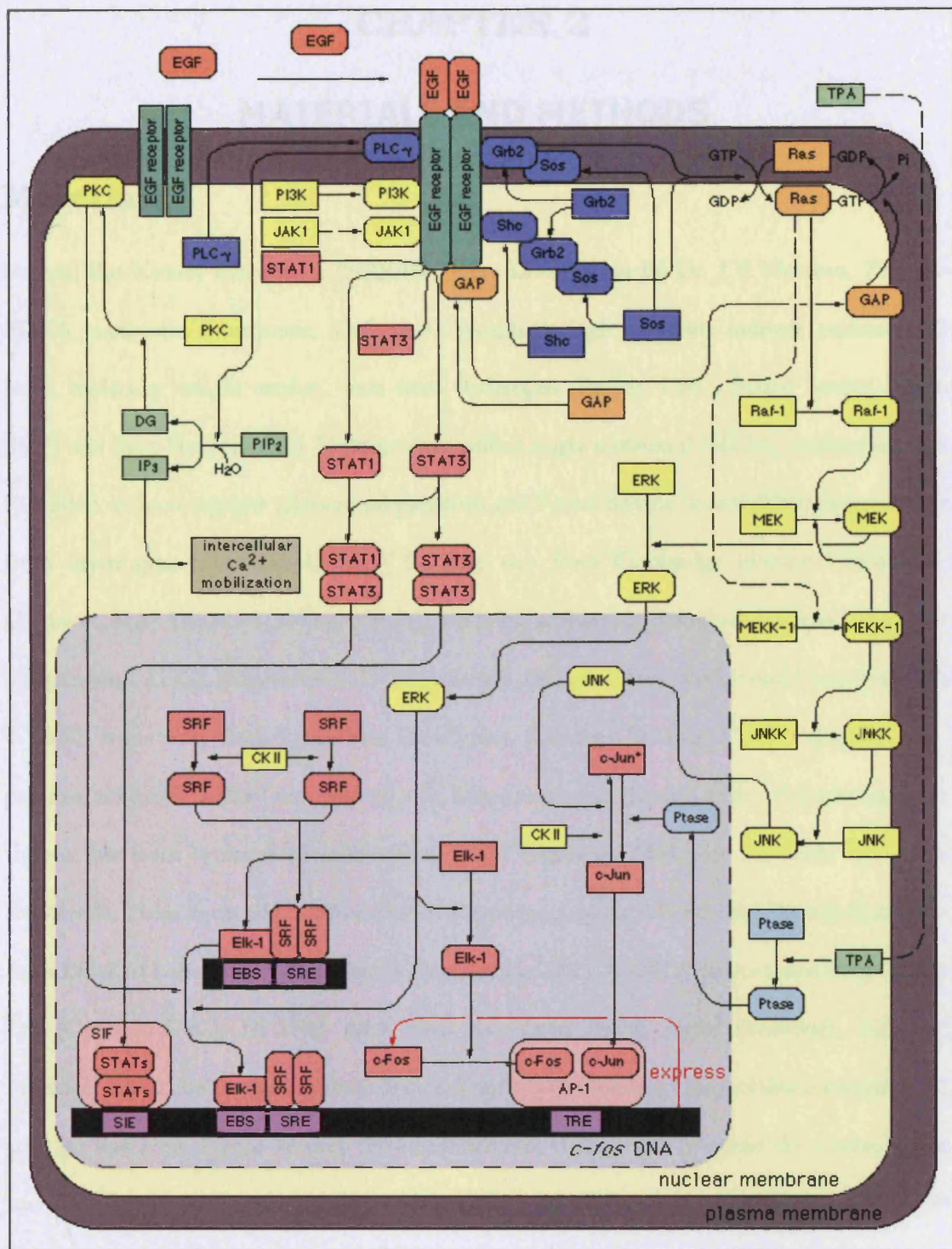
Although the EGFR induced pathways modules run in parallel (Figure 5), there is a considerable degree of cross-talk between them, which creates multiple opportunities for modulating or fine-tuning responses to different signals. MAPK cascades can influence other signalling pathways and vice versa; there are several known examples of MAPK influences the

other signalling pathways. For example, ERKs can activate the JAK-STAT pathways (Marshall 1995). MAPKs can also converge on the same substrates, like ERK and p38 which both activate MNK1 (Fukunaga et al. 1997), the MAPK signalling-integrating kinase 1, which is responsible for the activation of the eukaryotic translation initiation factor 4E, as well as MSK1.

Signalling pathways linked to cAMP and calcium are also influenced MAPK signalling cascades. The MEKs that activate ERKs can be activated by specific Raf isoforms based on the type of cell stimulated and the level of cAMP generated. Calcium signalling also affects ERK modules; it has been shown in cardiomyocytes and neurons that calcium influx results in MEF-2 activation, a transcription factor that is a substrate for the calcium/calmodulin-dependent protein kinase IV (Lu et al. 2000; Mao et al. 1999). The activation of MEF-2 correlates with the activation of p38, which is responsible for the phosphorylation of MEF-2 on its activation site.

A summary including different pathways branching from EGFR and indicating some of the crosstalks between pathways is reported in Figure 10.





**Figure 10. Scheme of different pathways involved in EGF signalling**  
(Taken from <http://www.grt.kyushu-u.ac.jp/spad/pathway/egf.html>).

## CHAPTER 2

### MATERIALS AND METHODS

#### Materials

Normal Rat Kidney Fibroblasts (NRK49F) were kindly given by Dr. J.T. Norman. Trypsin-EDTA, antibiotic/antimycotic, Dulbecco's modified Eagle medium nutrient mixture F12 Ham, molecular weight marker, were from Invitrogen (Paisley, U.K.). Foetal Bovine Serum (FBS) was from HyClone (lot). Dulbecco's modified Eagle medium (DMEM), nutrient mixture F12 Ham without arginine (custom preparation) and Foetal Bovine Serum (FBS) dialysed were from Invitrogen. L-arginine U- $^{13}\text{C}_6$  ( $^{13}\text{C}_6$  Arg) was from Cambridge Isotope Laboratories (Andover, MA). Duracryl (30% acrylamide, 0.8% bis acrylamide) was from Genomic Solutions (Huntingdon, U.K.). Disposable NAP-10 columns, ECLplus detection kit and 13cm linear pH 4-7 IPG strips were from Amersham Biosciences (Chalfont St.Giles, U.K.). Complete mini-protease inhibitor cocktail was from Roche Diagnostics Ltd (Lewes, U.K.). Sequencing grade trypsin was from Promega (Southampton, U.K.). Prestained SDS-page standards low range, acrylamide, 11cm linear pH 5-8 Immobiline DryStrips, Colloidal Gold total Protein Stain were from Bio-Rad Laboratories Ltd (Hemel Hempstead, U.K.). Ni-NTA agarose was from Qiagen Ltd (Crawley, U.K.), Ni-TED silica resin was from Active motif (Rixensart, Belgium), Fractogel EMD chelate (S) was from Merck (Darmstadt, Germany), immobilised iminodiacetic acid gel was from Perbio Science (Northumberland, U.K.) and Toyopearl AF-Chelate-650M was from Tosoh Bioscience (Stuttgart, Germany). Phospho-tyrosine monoclonal antibody (P-Tyr102), phospho-(Ser/Thr) PKA substrate antibody, Phospho-(Thr) MAPK/CDK Substrate monoclonal Antibody, Phospho-(Ser) PKC Substrate Antibody, phospho-p44/42 MAPK (Thr202/Thr204) E10 monoclonal antibody and p44/42 MAPK antibody were from Cell Signaling Technology (distributed by New England Biolabs, Hitchin, U.K.). Goat anti-mouse

or goat anti-rabbit Alexa-labelled secondary antibodies were from Molecular Probes, Oregon. All other chemicals were purchased from Sigma-Aldrich Company Ltd. (Poole, U.K.) and were of the best grade available.

## **Cell Culture**

### **Culture of NRK49F**

NRK49F cells were grown in p100 dishes in DMEM Nutrient mixture F12 Ham medium supplemented with 10% FBS plus antibiotics/antimycotic (100X solution contains 10,000U/ml penicillin-G, 10mg/ml streptomycin and 25l/ml amphotericin-B) at 37°C and 5% CO<sub>2</sub>. Cells were split when they were 70-80% confluent. Splitting involved washing the cells with 10ml PBS, trypsinising them with 1ml of Trypsin-EDTA and leaving them to incubate at 37 °C for 2 minutes. Detachment of cells from the dishes could be seen by naked eye. After detaching, cells were resuspended in fresh medium and plated. For stable isotopic labelling of amino acids in cell culture (SILAC), L-arginine and L-arginine U-<sup>13</sup>C<sub>6</sub> were added as to two separate lots of the arginine deficient media, to reach 147.5 mg/ml. NRK49F cell line was grown separately in the modified medium DMEM Nutrient mixture F12 Ham and normal media, supplemented with 10% dialysed FBS plus antibiotics at 37°C and 5% CO<sub>2</sub>. Cell lines were grown for at least eight splitting passages in media containing either <sup>12</sup>C<sub>6</sub> Arg or <sup>13</sup>C<sub>6</sub> Arg, before the preparation of cellular extracts.

### **Freezing of NRK49F cells**

NRK49F cells were grown until 70-80% confluent. They were trypsinised as mentioned above, resuspended in a small volume of medium and spun at 4°C for 5 minutes at 1000rpm. Supernatant was aspirated and cell pellet resuspended in freezing solution: 10% DMSO in

FBS. Cells were first left for 1-2 days (maximum 7 days) at -80°C, and then transferred to a liquid nitrogen container.

## **EGF stimulation of NRK49F cells**

Cells were grown until 70-80% confluence, and then the medium was replaced with the medium containing 0.5% FBS (starvation medium) and the cells were starved for 48h. On the day of the experiment, medium was replaced with fresh starvation medium 2 hours prior to stimulation. Cells were treated with EGF (200ng/ml) for 2, 5, 10 and 20 minutes at 37°C. After incubation dishes were placed on ice, washed twice with 5ml of ice-cold PBS and lysed in lysis buffer (50mM NaMOPS pH 6.8, 4% Zwittergent 3-12, 2% Triton X-100, 5mM NaF, 5mM Na Glycerolphosphate, 1mM  $\text{Na}_3\text{VO}_4$ , 5mM  $\text{Na}_4$  EDTA, 1mM  $\text{Na}_2\text{P}_2\text{O}_7$ , 1 tablet/10ml of complete mini protease inhibitor cocktail) .

## **Immunofluorescence Staining**

Cells were fixed in 4% paraformaldehyde in PBS for 15 min, quenched, and permeabilised in a solution of 50mM ammonium chloride and 0.2% saponin in PBS for 15 min, incubated with primary antibody 1:1000 in PBS supplemented with 1% gelatin and 0.02% saponin for 1 h, and then incubated with Goat anti-mouse or goat anti-rabbit Alexa-labelled secondary antibodies at 1:200 for 40 min. Confocal images were obtained using an MRC 1024 laser scanner (Bio-Rad, Hercules, CA) attached to an Optiphot 2 microscope (Nikon, Garden City, NY). Images were collated using Adobe Photoshop (Adobe Systems, Mountain View, CA).

## **In-vivo $^{32}\text{P}$ cells labelling**

NRK49F cells were incubated 14 hours in 0.25mCi/ml dissolved in phosphate free DMEM medium and 10% dialysed serum. Cells were washed with cold TBS containing orthovanadate 100mM.

## **IMAC purification of phosphoproteins**

Cell culture dishes were placed on ice, washed twice with 5ml of ice-cold PBS and lysed in 50mM NaMOPS pH 6.8, 4% Zwittergent 3-12, 2% Triton-X100, 1mM  $\text{Na}_3\text{VO}_4$ , 5mM NaF, 5mM  $\beta$ -glycerophosphate, 5mM EDTA and protease inhibitor. The cellular suspension was sonicated for 15 seconds three times and spun for 20 minutes at 13000rpm at 4°C. 8.0M LiCl was added to the supernatant to a final concentration of 2M LiCl. One part of bind mix (2M LiCl in lysis buffer containing 20% silica) was then mixed to 6 parts of protein solution and incubated on a rocking platform at room temperature for 5 minutes. The suspension was then centrifuged for 3 minutes at 13000 rpm and the supernatant was saved, the silica discarded. The protein solution was then rebuffered into binding buffer (50mM bis-(hydroxyethyl)-piperazine pH 3.4, 4% Zwittergent 3-12, 2% Triton X-100) using NAP-10 size exclusion columns according to manufacturer's instructions. For each  $\text{Fe}^{3+}$  affinity column in the 1ml of 50% slurry of Ni-NTA Agarose beads was pipetted into 2.5ml plastic columns.  $\text{Ni}^{2+}$  ions were removed from the resin by washing the column with 4ml of regeneration buffer (50mM Tris-HCl pH 7.5, 2% Triton X-100, 0.1M EDTA, 0.5M NaCl). The column was then washed with 4ml of ddH<sub>2</sub>O to remove any trace of EDTA and subsequently activated with 2ml of freshly dissolved  $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$  (27mg/ml in water). The column was then washed with 4ml of water and equilibrated with 4ml of binding buffer. 1.5ml of protein solution containing usually between 1-3 mg of protein were loaded on the  $\text{Fe}^{3+}$  affinity column: at all times the column was run by gravity force alone. After loading the sample the columns were washed with 3ml of binding buffer and then with 1ml of 25mM  $\text{Na}_2\text{SO}_4$  and 25mM imidazole in binding buffer. Phosphoproteins were eluted in 1ml of 50mM  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$  in binding buffer. The columns, after extensive washing with binding buffer and water, can be recycled starting with the regeneration buffer step. Phosphoproteins were precipitated with 7.5% TCA and 0.02% 1, 10 phenantroline on ice over night. They were then spun for 20 minutes at 13000 rpm at 4°C.

Supernatant was aspirated and the protein pellets were washed once with ice-cold 5% TCA for 10 minutes, and 4 times with ice-cold 80% acetone. Washing includes 5 minutes incubation on ice and subsequent 5 minutes spinning at 13000rpm at 4°C. After the last wash protein pellets were left on the bench to air-dry for a few minutes and then resuspended in isoelectric focusing buffer (8M urea, 2M thiourea, 4% CHAPS, 1% Triton X-100, 65mM DTT, 10mM Tris base, 0.8% ampholytes).

## **Protein assays**

### **Bicinchoninic acid protein assay**

This is a modification of a method originally proposed by (Smith et al. 1985). The principle of the method is that BCA reduces divalent copper ion to monovalent ion in alkaline conditions. A molybdenum/tungsten purple product is produced with an absorbance maximum at  $\lambda=562\text{nm}$ . The reagents used for the assay were Bicinchoninic acid solution and Copper (II) sulphate pentahydrate 4% solution. A standard curve was prepared with BSA. BSA concentrations were chosen in the range between 2 $\mu\text{g}$  to 100 $\mu\text{g}$ , depending on the approximate estimation of the concentration of the unknown proteins.

BCA protein assays were performed in microtiter 96-well plates. The working BCA reagent solution was prepared by mixing 1 part of Copper (II) sulphate pentahydrate 4% solution to 50 parts Bicinchoninic acid solution. The working solution was added to the BSA and unknown protein sample in ratio 20:1. The plate was incubated for 30 minutes at 37°C. After incubation the plate was cooled to room temperature and absorbance measured at  $\lambda=595\text{nm}$ . Control for each set of samples was the corresponding buffer. The absorbance of the control was subtracted from the absorbance of the samples to obtain the net absorbance of the proteins. A standard curve was plotted as the net absorbance at  $\lambda=595\text{nm}$  vs. the known  $\mu\text{g}$  of BSA. This

curve was used to determine the amount of protein in unknown samples. This method is not compatible with reducing agents such as DTT.

## **Modified Bradford protein assay**

This method is property of ProteoSys AG, Germany.

## **Electrophoresis of proteins**

### **One-dimensional SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot**

This electrophoresis was based on the protocol originally proposed by Laemmli (Laemmli 1970). Acrylamide used for all gels was a mixture of 30% Acrylamide and 0.8% Bisacrylamide. Separating gels were prepared in the range from 10% - 12% depending on the size of proteins. Separating gel was prepared with different volumes of 30% Acrylamide and 0.8% Bisacrylamide, 0.375M Tris HCl pH 8.8, 0.1% SDS and APS and TEMED were used for polymerisation. All stacking gels were of 5% acrylamide concentration and contained 0.125M Tris HCl pH6.8, 0.1% SDS. Sizes of gels were either 15cm x 16cm (Hoefer SE600 system, Amersham Pharmacia) or 8.3cm x 7.3cm (Mini-Protean 3 system, Bio-Rad). Samples were prepared in SDS-PAGE sample buffer (5X solution: TrisHCl pH 6.8 0.225M, glycerol 50%, SDS 5%, bromophenol blue 0.05%, DTT 0.25M), vortexed and heated for 5 minutes at 95°C.

Proteins from cellular extracts and proteins obtained after phosphoprotein enrichment were separated on 10% SDS-PAGE gels, and then transferred onto nitrocellulose membranes by semi-dry blotting. Transfer was performed at 21V for 90minutes. Membranes were blocked with 7% non-fat dry milk or 5% BSA in TBS/Tween buffer (20mM Tris-HCl, pH 7.4, 150mM NaCl, 0.1% Tween 20) for minimum 1 hour at room temperature (or overnight at 4°C). After

blocking, membranes were washed quickly with TBS/Tween and incubated with the primary antibodies for a minimum of 1 hour at room temperature (or overnight at 4°C). Primary antibodies were diluted in TBS/Tween according to the manufacturer's instructions. Membranes were then washed 3 times with 150ml TBS/Tween for 10 minutes and incubated with secondary antibodies for 1 hour at room temperature. Horseradish peroxidase-conjugated secondary antibodies were used at a 1:10000 dilution in TBS/Tween. After incubation membranes were washed 3 times for 10 minutes with TBS/Tween and then once with TBS for 5 minutes. Detection was done using the ECLplus detection kit.

Antibodies were stripped off membranes by incubating in stripping buffer (62.5mM Tris-HCl, pH 6.8, 100mM  $\beta$ -mercaptoethanol and 2% SDS) for 30 minutes at 60°C under agitation. Membranes were then washed with TBS/Tween for 3 times for 10 minutes and reblocked as described above.

After a maximum of 5 different antibodies were tested the membrane was washed with 100ml of TBS/Tween 3 times for 20 minutes, rinsed with 100ml of ddH<sub>2</sub>O for 3 times for 5 minutes, and incubated in the minimal amount of colloidal gold to cover the surface until the staining was satisfactory (4-5 hours). The membrane was then drained and let dry overnight to increase the signal to noise ratio.

Western blot films were scanned using ImageScanner (Amersham Pharmacia Biotech) and the image analysed with ImageQuant software (Amersham Biosciences).

## **Two-dimensional- polyacrylamide gel electrophoresis (2D-PAGE)**

13cm linear pH 4-7 IPG strips were loaded with 150 $\mu$ g of proteins by passive rehydration overnight. Isoelectric focusing was performed on an IPGphor device (Pharmacia Biotech) and was carried out up to a total of 66000 Volt-hours. The IPG strips were rinsed thoroughly with



distilled water, quickly dried on filter paper and focused proteins were reduced (50mM Tris/HCl, pH 6.8, 6M urea, 2% w/v SDS, 30 % v/v glycerol, 2% w/v DTT) and alkylated (50mM Tris/HCl, pH 6.8, 6M urea, 2% w/v SDS, 30 % v/v glycerol, 4.5% w/v iododacetamide) for 20 minutes each. IPG strips were rinsed thoroughly with distilled water, and the excess water was drained onto a piece of filter paper. Gels (15cm x 16cm x 1mm) of varied acrylamide concentration 10% - 12% were used, depending on the size of the proteins being separated.

Molecular weight markers were loaded onto a piece of filter paper and placed close to the acidic part of the IPG strip onto the polyacrylamide gel. The strip and the markers were overlaid with 0.5% boiling agarose prepared in stacking gel buffer with a trace of bromophenol blue. Electrophoresis ran at a constant voltage of 200V for approximately 5 hours at 11°C. When the bromophenol blue dye reached the end of the gel the run was stopped and the gel was further stained with either silver (Shevchenko et al. 1996) or Coomassie brilliant blue R250.

## **Silver staining of polyacrylamide gels**

This procedure was adopted from Shevchenko (1996). Gels were placed in fixing solution (12%  $\text{CH}_3\text{COOH}$  and 50% methanol) and agitated for a minimum of 2 hours. They were then washed 3 times for 20 minutes with 50% ethanol. The gels were then pre-treated for 1 minute with 0.02%  $\text{Na}_2\text{S}_2\text{O}_3$  and rinsed well with distilled water 3 times for 20 seconds each time. The staining step was performed in 0.1%  $\text{AgNO}_3$  for 20 minutes. Excess silver nitrate was washed off twice with water, and the proteins spots were detected with a solution containing 3%  $\text{Na}_2\text{CO}_3$ , 0.0002%  $\text{Na}_2\text{S}_2\text{O}_3$  and 0.009% formaldehyde. When the spots became a dark brown colour, the gels were rinsed twice for 2 minutes in distilled water. Staining was stopped by immersing the gels in the fixing solution for 10 minutes. After this the gels were washed in

50% methanol, and finally in 5% methanol. Gels were then packed in plastic wraps with a small amount of 5% methanol, to stop them from drying and they were stored at 4°C until the proteins were cut and analysed by mass spectrometry.

## **In-gel tryptic protein digestion**

Protein spots of interest were excised from the SDS gels and chopped into small pieces (1mm<sup>3</sup>). The gel spots were automatically processed using Genetic Solution Processor or manually as described here. Gel pieces were treated with 25mM NH<sub>4</sub>HCO<sub>3</sub> for 15 minutes. This step was repeated until the pH of the solution was alkaline. Protein spots were then washed three times with acetonitrile and dried in the speed-vacuum concentrator for 30 minutes. The dry gel clump was rehydrated in 10mM DTT, 100mM NH<sub>4</sub>HCO<sub>3</sub> for 30 minutes then spun down and the supernatant removed. The gel was washed again in acetonitrile and dried in the speed-vacuum concentrator for 5 minutes. Trypsin solution was prepared by dissolving 20µg of trypsin in 500µl of 25mM NH<sub>4</sub>HCO<sub>3</sub> in ice. 5µl of this solution was added to each sample, left to rehydrate for a while at 4°C, and then if necessary more trypsin solution was added just to cover the gel, without letting it dry. The final concentration was kept between 6-30 ng/µl. These samples were then incubated overnight at 30°C.

## **Mass Spectrometric Analysis**

### **MALDI-TOF Mass Spectrometry**

For MALDI-TOF mass spectrometry aliquots of 0.5µl of the digest solution were applied to a target plate and allowed to air-dry. Subsequently, 0.5µl of matrix solution (1% w/v α-cyano-4-hydroxycinnamic acid in 33% acetonitrile and 0.1% v/v trifluoroacetic acid) was applied to the dried sample and the sample was again allowed to dry. Spectra were obtained using Bruker Biflex III (Bremen, Germany) spectrometer in reflector mode. The spectrometer was calibrated

using a mixture of 5 known peptides ranging in mass between 842 and 3495 Daltons. Obtained spectra were processed manually using Bruker's DataAnalysis software, version 1.6.

## **Mass Fingerprinting (MFP)**

Trypsin is known to cleave proteins after a lysine or arginine residue. Therefore if the sequence of a protein is known, it is possible to predict what masses would be produced if that protein were cleaved by trypsin. Several internet-based databases now exist in which the theoretical digests of known proteins are stored. By matching the spectrum of masses obtained by mass spectrometry with these databases it is possible to identify the starting protein from a tryptic digest sample. This technique is known as “mass fingerprinting” (see introduction). Searching the database provides a list of possible matches and a score for each indicating the probability that that protein is the correct match. The larger the number of peptides compared, and the more accurate the mass measurements made, the more likely it is that this method of protein identification will be successful. However, possible matches may have similar probabilities, and factors such as post-translational modification may interfere with database matching. If a clear match can not be found, or the location of a post-translational modification is to be pinpointed, the remaining sample can be sequenced using an ESI-Ion Trap mass spectrometer.

Protein fragments were identified using PROFOUND programme available at Prowl website ([http://prowl.rockefeller.edu/profound\\_bin/WebProFound.exe](http://prowl.rockefeller.edu/profound_bin/WebProFound.exe)). Mass finger printing of SILAC sample were all conducted allowing a partial modification on arginine of +6 Dalton.

## **ESI-Ion Trap Mass Spectrometry**

Tryptic digest samples were almost completely dried and resuspended in 50% methanol and 5% formic acid. Nanospray mass spectrometry was performed using a Finnigan-Matt LCQ (San Jose, CA, USA), in a positive ion mode. Collision energy changed between 28 and 35%.

Peaks previously identified using MALDI-TOF MS were chosen for sequencing, both to confirm the identity of the protein, and the phosphorylation status of the tryptic peptides. Mass values corresponding to high probability phosphorylated peptides were also analysed. The probability of the modification on specific residues was determined using scansite (<http://scansite.mit.edu/motifscan>) and prosite (<http://ca.expasy.org/prosite/>). Sequences were scanned for specific motifs using patinprot software (available on [www.expasy.org](http://www.expasy.org)). Fragmentation spectra were compared to theoretical fragmentations of the likely sequences obtained from MS-Product (<http://jspl.ludwig.edu.au>), and the location of post-translational modifications ascertained where possible.

## **CHAPTER 3**

### **PHOSPHOPROTEIN ENRICHMENT**

#### **INTRODUCTION**

Protein phosphorylation is a fundamental process for regulation of the activity of proteins in signal transduction and in a variety of cellular functions including metabolism and apoptosis (Hunter 1995; Pawson et al. 2003). One third of the proteins expressed in mammalian cells are phosphorylated at serine, threonine and, less commonly, tyrosine residues (Krebs 1994), hence protein phosphorylation is considered to be one of the most important posttranslational modifications (Annan et al. 1996; Neubauer et al. 1999). This justifies an intense development of methods to detect, analyse and quantify phosphoproteins.

On one hand, focus has been given to staining procedures specific for phosphoprotein (Ge et al. 2004; Patton 2002; Schulenberg et al. 2004). On the other hand, phosphoproteins separated on gels have been followed by either radiography after  $^{32}\text{P}$  radiolabelling (Immler et al. 1998; Kang et al. 1997), or western blotting with phospho specific antibodies (Soskic et al. 1999; Stannard et al. 2003a; Steen et al. 2002). Detection of  $^{32}\text{P}$ -labelled phosphoproteins is very sensitive, but it is only applicable to cells in culture. Since it is much more sensitive than silver staining, low yields in subsequent protein identifications have been a limitation of the overall procedure (Adam et al. 2002). In any case, one of the most powerful techniques for separation of phosphoproteins continues to be gel electrophoresis.

Identification of the proteins and assignment of the phosphorylation sites is generally achieved by MS, although phosphopeptides are often of low abundance and difficult to ionise. This complicates their analysis in very complex peptide mixtures and it has initially led to the

development of procedures for extraction and enrichment of phosphopeptides based immobilised metal ion affinity chromatography (IMAC).

This separation technique uses solid chromatographic supports with covalently bound chelating compounds that complex metal ions and serve as affinity ligands for nucleophilic groups in peptides/proteins. Metal ions can be divided into three categories based on their reactivity towards nucleophiles, namely hard, intermediate and soft. Hard metal ions like  $\text{Fe}^{3+}$ ,  $\text{Ca}^{2+}$ , and  $\text{Al}^{3+}$  show a preference for oxygen, soft metal ions such as  $\text{Cu}^+$ ,  $\text{Hg}^{2+}$ , and  $\text{Ag}^+$  prefer sulfur and intermediate metal ions ( $\text{Cu}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Co}^{2+}$ ) coordinate nitrogen, oxygen and sulfur. High affinity of Fe(III) for the oxygen atoms in the phosphorylated side chains of serine, threonine and tyrosine, as well as carboxylic acids, has been exploited to isolate phosphopeptides under acidic conditions (Chaga 2001). Substantial progress has been made with this approach and its application as a reliable method for high-throughput analysis of phosphorylated peptides by mass spectrometry (Jin et al. 2004; Liu et al. 2005; Nuhse et al. 2003). One of the main problems encountered was protein solubility under the acidic conditions required in the isolation procedures.

Previous studies have underlined difficulties in separating phosphoproteins from proteins with surface exposed clusters of carboxylic groups using IMAC. The affinity of phosphate groups to immobilised Fe(III) and its use in isolating phosphoproteins was first described by Porath and colleagues in 1986 (Andersson et al. 1986; Muszynska et al. 1992). In these early studies, all 20 naturally occurring amino acids plus the phosphorylated forms of serine, threonine and tyrosine were tested for affinity to Fe(III) immobilised on IDA-sepharose under different pH conditions. The phosphoamino acids showed tight binding at pH 3.1 and 5.0, but no affinity at pH 7.2 and 8.5. L-aspartic acid and L-glutamic acid were the only other amino acids that showed some affinity, although very weak. In a later study the same group demonstrated that,

although the interaction between carboxyl groups and chelated Fe(III) ions is very weak compared with the phosphate-Fe(III) interactions, when there are clusters of carboxylic and phenolic groups, the chelating effect is increased, becoming almost comparable to that for phosphate groups.

Current knowledge about cellular biology suggests that a more comprehensive analysis of the role of phosphorylation in signalling can be obtained by isolation and analysis of phosphoproteins rather than phosphopeptides (Godovac-Zimmermann et al. 2003). The majority of genes in the human genome encode for polypeptides that become co- and posttranslationally modified. Many different isoforms of a single protein are produced during different cellular responses, at different times and in different cell types. In a single cell, a single protein can have many different isoforms which are involved in different functional processes. The roles and partitioning of these isoforms between different functions are extremely difficult to study with phosphopeptide methods since protein fragmentation by tryptic hydrolysis irretrievably destroys and superimposes the functionally crucial information on isoform distributions (Stannard et al. 2004). Furthermore, efficiencies and patterns of tryptic fragmentation of proteins are known to be changed by post-translational modifications such as phosphorylation and this complicates determination of accurate, quantitative relationships between phosphorylation and cellular function. The interest has therefore been in monitoring phosphoproteins separately and in identifying the roles of specific phosphorylation isoforms in different cellular states or tissues (Godovac-Zimmermann et al. 2003).

In this chapter the development of an efficient protocol to strongly enrich phosphoproteins from complex mixtures based on home made, recyclable Fe(III)-affinity columns and inexpensive buffers is described. In the following chapters different applications of this optimised method to study EGF signalling are exposed.

## RESULTS

### Development of phosphoprotein enrichment procedure

A scheme of the phosphoprotein extraction protocol is shown in Figure 11. Cells were lysed in the presence of protease and phosphatase inhibitors under high detergent concentration in a 3-(N-Morpholino) propane-sulfonic acid buffer at pH 6.8. As the solubility of phosphoproteins is strongly dependent on pH, a key feature of this procedure was the use of different pHs for recovery of total proteins from cell lysates (pH 6.8) and for subsequent capture of phosphorylated proteins (pH 3.4). In the course of the method development the ability of many different combinations of detergents and detergent concentrations to retain proteins in solution at pH 6.8 and at pH 3.4 was tested. Zwittergent detergents (Zwittergent 3-8, 3-10, 3-12 and 3-14) were preferred because, unlike other amphoteric surfactants, they retain their zwitterionic character over a wide pH range. Their ionic balance prevents irreversible binding of the detergent to either anionic or cationic compounds. The most effective combination proved to be 4% Zwittergent 3-12 together with 2% Triton X-100.

The following step is a simple DNA removal which exploits the affinity of DNA for silica in the presence of high ionic strength. This DNA removal step is strictly necessary since the DNA was otherwise absorbed to the affinity matrix via its phosphate backbone. For the amounts of cell lysates typically processed, a minimum of 0.2 mg/ml of silica was necessary to ensure complete DNA removal in the presence of 2M LiCl (data not shown).

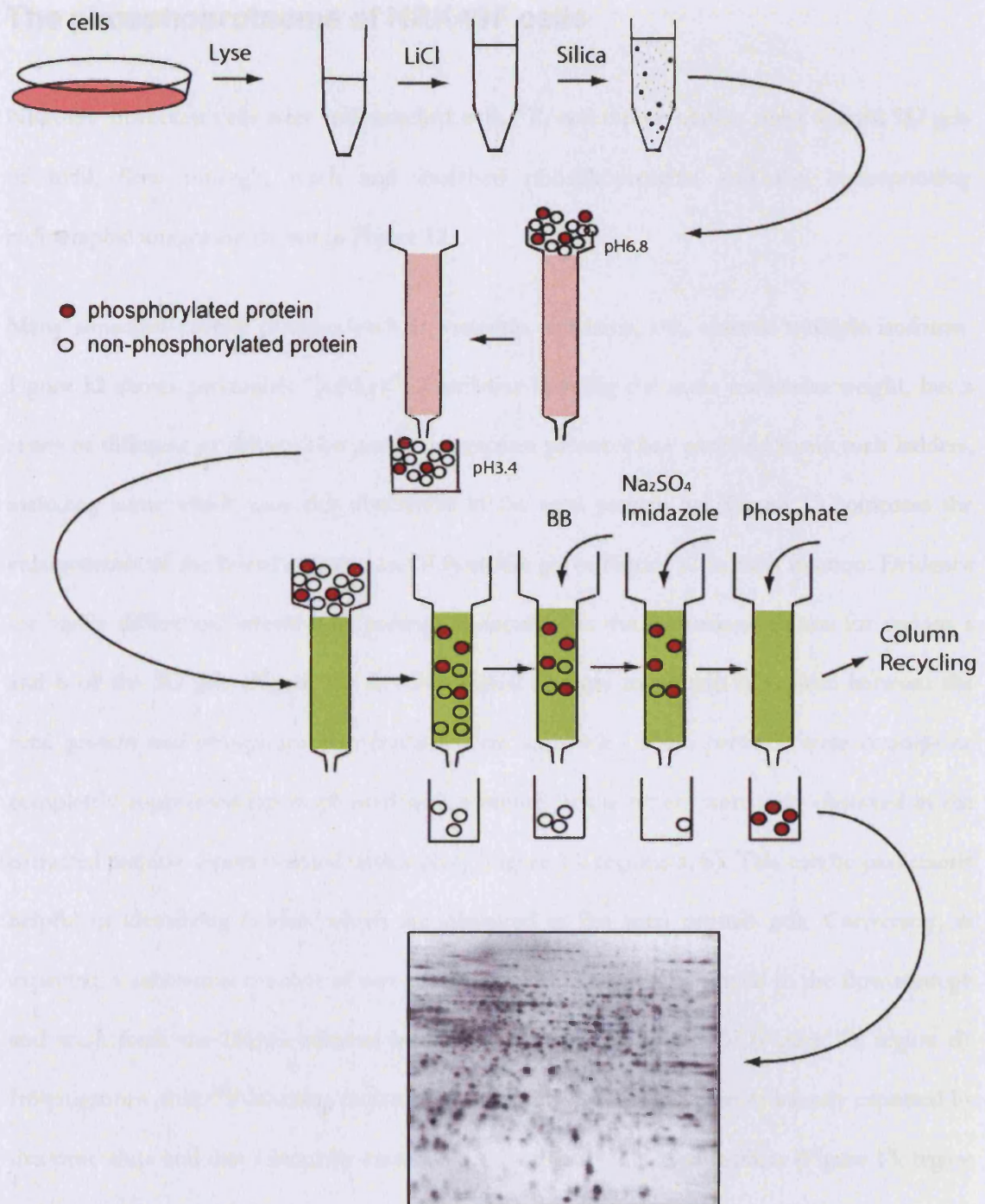
A low pH value (pH 3.4) is expected to maximise the affinity and specificity of the immobilised Fe(III) for phosphate groups. The protein lysate was therefore exchanged with a N,N'-Bis(2-hydroxyethyl)piperazine buffer containing the same detergents (binding buffer). The Fe(III) affinity columns were prepared just prior to use by treating the chelating resin with a FeCl<sub>3</sub> solution and then equilibrating the affinity column with the low pH binding buffer.



A variety of different immobilised Fe(III) media have also been tested (see discussion). Amongst these media, the use of NTA-agarose was preferred due to its good performance with the general conditions used here, its reported superiority over IDA-agarose in the isolation of phosphopeptides (Zhou et al. 2000) and potential advantages in the mode of binding of phosphate groups to Fe(III) as compared to carboxylate groups (Muszynska et al. 1986; Muszynska et al. 1992) (see discussion).

Following absorption of the phosphoproteins, the columns were washed to reduce non-specific binding. A first general wash was performed with binding buffer and subsequently the columns were washed with binding buffer containing 25mM Na<sub>2</sub>SO<sub>4</sub> and 25mM imidazole (pH 3.4) to minimise interaction of histidine groups with the immobilised Fe(III). For 0.5ml columns, loading of increasing amounts of protein lysates (1.5 - 12 mg) gave a constant proportion between protein loaded and protein eluted up to 3 mg (data not shown). This capacity is consistent with capacities reported for commercially available columns.

The absorbed phosphoproteins were eluted using 50mM sodium phosphate in binding buffer (pH 3.4). The elution efficiency was checked by varying the phosphate concentration and the pH value of the buffer (data not shown). The elution of the phosphoproteins was primarily dependent on increasing phosphate concentration rather than on increasing pH value, as has been found by others for phosphopeptides (Hart et al. 2002). Yields of phosphoproteins relative to total proteins were typically between 10-13% for NRK49F cells, 17-23% for HeLa cells, and 20-25% for rat liver tissue.



**Figure 11. Schematic representation of the protocol used for IMAC extraction of phosphoproteins**

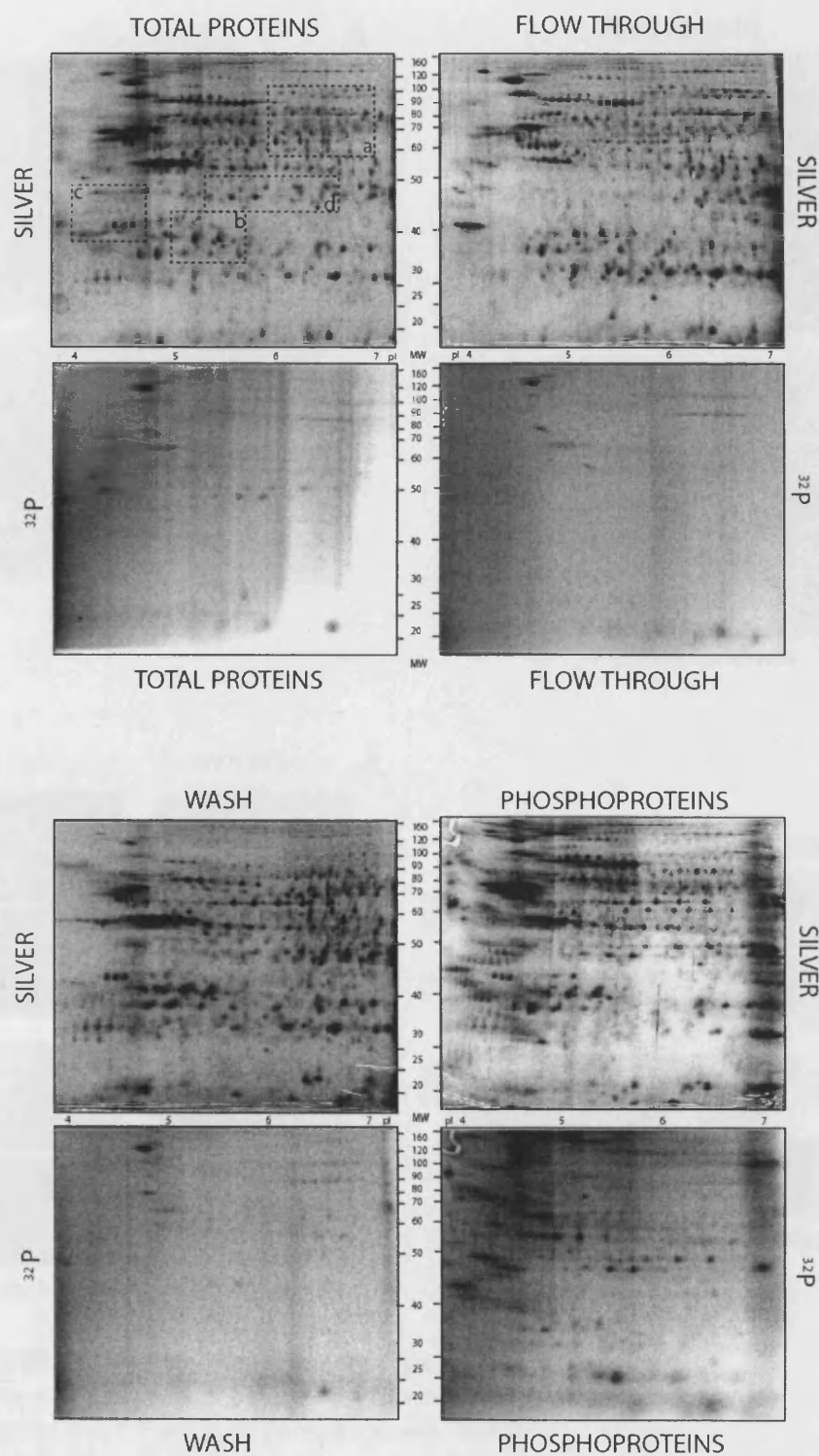
Cells were lysed and the LiCl was added. The lysate was then treated with silica for DNA removal. The total proteins were rebuffed in a high detergent buffer at pH 3.4. The proteins were then loaded on IMAC column and, after one non-specific and one specific wash, the phosphoproteins were eluted with a phosphate buffer at the same pH. Proteins were then TCA precipitated and resuspended in IEF buffer for 2D electrophoresis. 2D gel is of phosphoproteins extracted from NRK49F cells over pI 5-8.

## The phosphoproteome of NRK49F cells

NRK49F fibroblast cells were radiolabelled with  $^{32}\text{P}$ , and representative silver stained 2D gels of total, flow through, wash and enriched phosphoproteins and the corresponding radiographic images are shown in Figure 12.

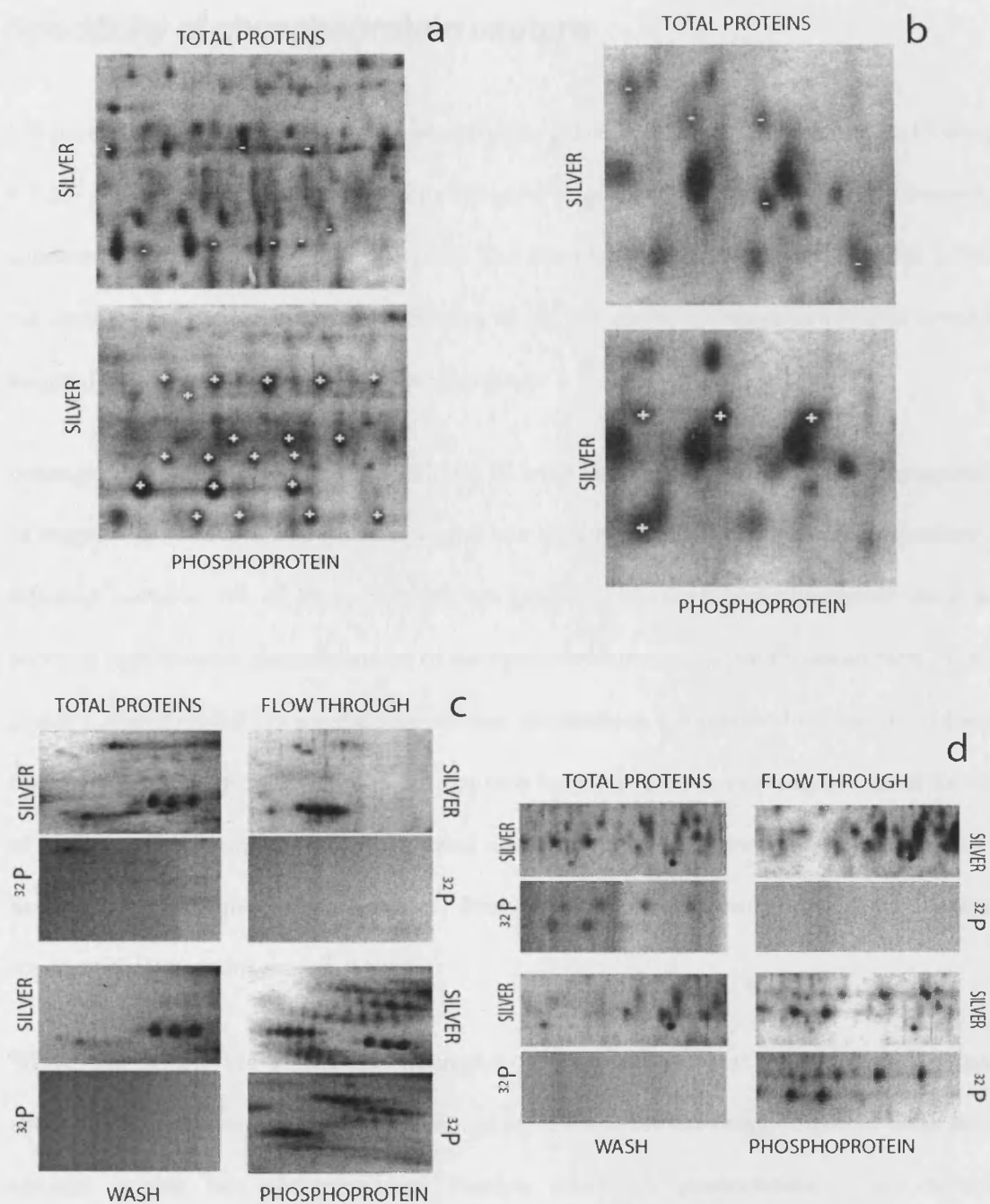
Many abundant cellular proteins such as vimentin, annexins, etc., exist in multiple isoforms. Figure 12 shows prominent “ladders” of proteins showing the same molecular weight, but a series of different pI values. The present extraction protocol has enriched many such ladders, including some which were not observable in the total protein gel. Figure 13 compares the enlargements of the boxed region c and d from the gel in Figure 12 in each fraction. Evidence for highly differential selection of proteins is apparent in the expansions shown for regions a and b of the 2D gels (Figure 13, ab). Substantial changes in protein intensities between the total protein and phosphoprotein fraction were apparent - some proteins were strongly or completely suppressed (spots pointed with a minus), while others were only observed in the extracted material (spots pointed with a plus) (Figure 13, regions a, b). This can be particularly helpful in identifying ladders which are obscured in the total protein gels. Conversely, as expected, a substantial number of non radioactive proteins were observed in the flow-through and wash from the IMAC column, but not in the extracted material (Figure 13, region d). Investigations with  $^{32}\text{P}$  labelling indicated that  $^{32}\text{P}$  labelled proteins are efficiently captured by this procedure and that selectively enriched protein spots contain phosphate (Figure 13, region c and d).

Although complete removal of other proteins was not obtained for all proteins without  $^{32}\text{P}$  labelling (Figure 12), phosphoproteins could be reliably identified on the silver-stained gels by their selective enrichment in the extracted material.



**Figure 12. Silver stained 2D gels for total proteins, flow through, wash and phosphoproteins extracted from  $^{32}\text{P}$  radio labelled NRK49F cells**

Equal amount of total proteins, flow through, wash and phosphoprotein fractions were separated on 2D gels, using 13m pI 4-7 strips on 12% acrylamide gels, and the gels silver stained. For each gel the corresponding  $^{32}\text{P}$  image is reported.



**Figure 13. Phosphoprotein enrichments specificity**

Enlargements refer to gels in Figure 12. In a and b, - indicates spots lost in the chromatography and + spots of phosphoprotein enriched.

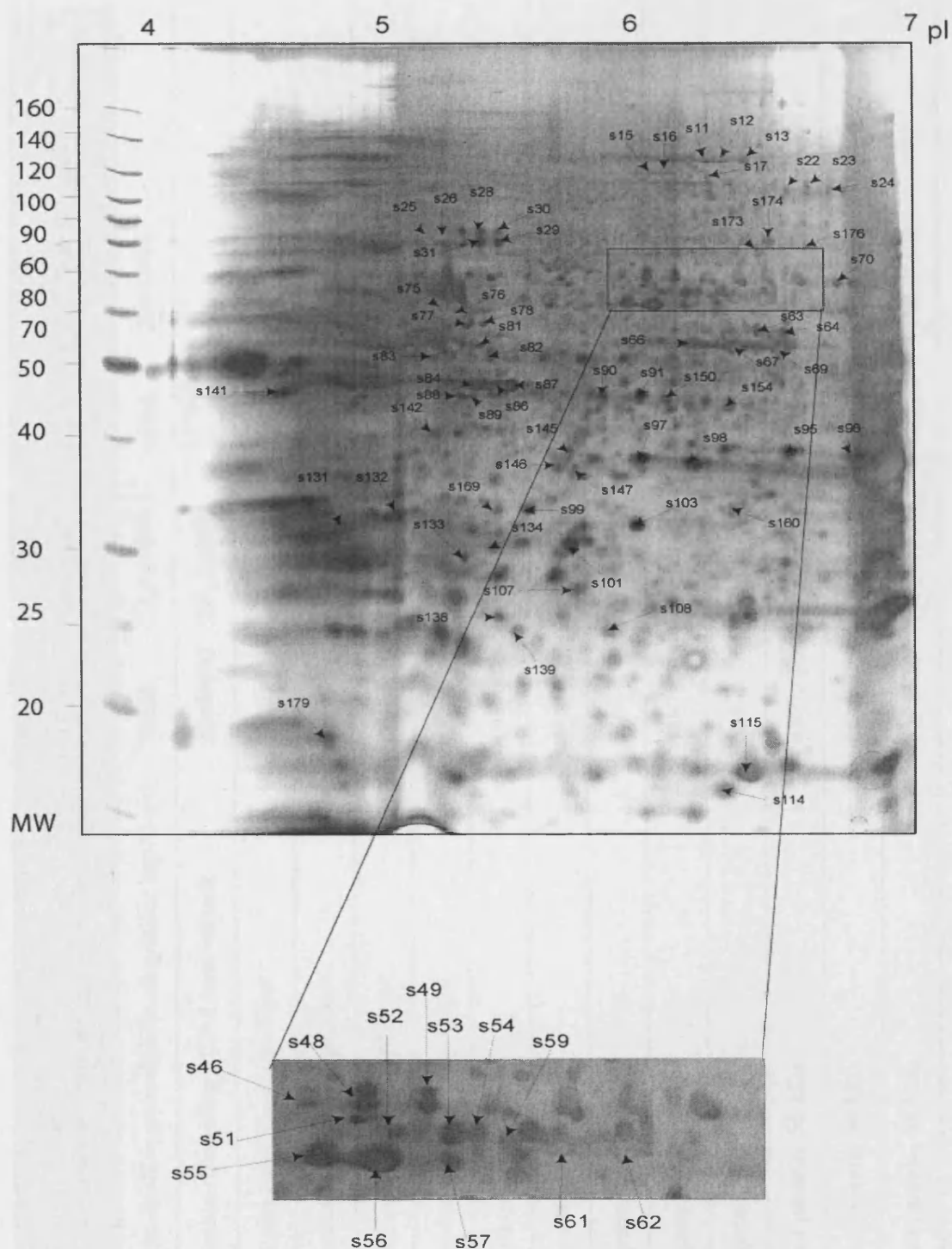
## **Specificity of phosphoprotein capture**

108 protein spots enriched in the phosphoproteins gel were randomly chosen from a 13 cm, pI 4-7 2D gel and the proteins subjected to MS mass fingerprinting (Figure 14) (data obtained in collaboration with Jelena Predic-Atkinson). The identified proteins are listed in Table 1. With the recent rapid progress in characterisation of the rat genome, almost all proteins could be assigned to rat proteins and most to specific genes.

Amongst the 108 spots which were analysed, 94 were identified (most of the rest corresponded to superposition of different proteins in the low resolution gels used here). They matched 60 different proteins. All of these proteins are predicted to have phosphorylation sites, and previous experimental demonstrations of phosphorylation are available for about 80% of them (Table 1, the PubMed ID numbers of relevant publications are reported on the last column). For many of the 51 proteins that have previously been reported to be phosphorylated the sites of phosphorylation are not known in detail and only some have previously been reported to have multiple phosphorylation isoforms. Between 2 and 5 different isoforms were identified for 16 of these proteins on the 2D gels.

While this is not yet exhaustive (examples of detailed analysis of the phosphorylation sites/isoforms for some individual proteins is reported in the following chapters), these results strongly suggest that phosphorylated proteins are highly predominant in the extracted phosphoprotein fraction.





**Figure 14. Phosphoprotein fraction separated on 2D electrophoresis gel**  
 The phosphoprotein fraction was separated on 2D PAGE using a 13cm pI 4-7 strip, on a 13% acrylamide gel. The spots labelled have been cut and identified using MALDI-TOF MS.

Spot	Proteins that act as chaperones/mediators of protein folding	GI	Accession	Gene	PMID Reference
43.1	Mixture: similar to dnaK-type molecular chaperone hsp72-ps1	56385	CAA49670.1	none	10843916
64	Similar to chaperonin containing TCP-1 beta subunit	34864883	XP_216891	LOC299809	
39	dnaK-type molecular chaperone hsp72-ps1	347019	S31716	none	10843914
35	dnaK-type molecular chaperone hsp72-ps1	347019	S31716	none	10843914
36	dnaK-type molecular chaperone hsp72-ps1	347019	S31716	none	10843914
37	dnaK-type molecular chaperone hsp72-ps1	347019	S31716	none	10843914
54	Endoplasmic oxidoreductase 1	19744821	AAL96669.1	none	
28	FK506 binding protein 10	34873708	XP_340902.1	Fkbp10	7493967
29	FK506 binding protein 10	34873708	XP_340902.1	Fkbp10	7493968
30	FK506 binding protein 10	34873708	XP_340902.1	Fkbp10	7493969
31	FK506 binding protein 10	34873708	XP_340902_1	FKBP10	7493969
55	Glucose regulated protein, 58 kDa	8393322	NP_059015.1	Grp58	10870974
56	Glucose regulated protein, 58 kDa	8393322	NP_059015.1	Grp58	10870975
57	Glucose regulated protein, 58 kDa	8393322	NP_059015.1	Grp58	10870976



26.2	Heat shock 70kD protein 5	25742763	NP_037215.1	Hspa5	7517674; 3141786
25	Heat shock 70kD protein 5	25742763	NP_037215.1	Hspa5	7517674; 3141786
38	Heat shock protein 8	38181549	AAH61547.1	HSPA8	12080070; 10600240
155	Heat shock protein 60 (liver)	11560024	NP_071565.1	Hsp60	9724719; 15252132
156	Heat shock protein 60 (liver)	11560024	NP_071565.1	Hsp60	9724719; 15252132
22	Procollagen lysine, 2-oxoglutarate 5-dioxygenase 2	28400783	CAD23630.1	Plod2	
23	Procollagen lysine, 2-oxoglutarate 5-dioxygenase 2	28400783	CAD23630.1	Plod2	
24	Procollagen lysine, 2-oxoglutarate 5-dioxygenase 2	28400783	CAD23630.1	Plod2	
26.1	Protein disulfide isomerase related protein (calcium- binding protein, intestinal-related)	16758712	NP_446301.1	Erp70	8631326; 10079081
63	Similar to chaperonin containing TCP-1 beta subunit	34864883	XP_216891	LOC299809	
70	Stress-induced-phosphoprotein 1 (Hsp70/Hsp90- organising protein)	20302113	NP_620266.1	Stip1	By similarity 14754904; 11154072
157	T-complex protein 1, theta subunit (TCP-1-theta) (CCT-theta)	34867525	XP_213673.2	Cct8	by similarity to HSP60
Spot	Protein that act as metabolic enzymes	Gi	Accession	Gene	PMID Reference
15	Alpha glucosidase 2	34861580	XP_215144.2	Ganab	8094613; 246728931; 59730
16	Alpha glucosidase 2	34861580	XP_215144.2	Ganab	8094613; 246728931; 59730

66	Enolase-1, alpha	6978809	NP_036686.1	Eno1	8578591; 1956339; 6330085
67	Enolase-1, alpha	6978809	NP_036686.1	Eno1	8578591; 1956339; 6330086
69	Enolase-1, alpha	6978809	NP_036686.1	Eno1	8578591;1956339; 6330087
103	Enolase-1, alpha	6978809	NP_036686.1	Eno1	8578591;1956339 6330085
83	Enolase-2, gamma	26023949	NP_647541.1	Eno2	8578591;1956339; 6330087
62	Glucose-6-phosphate dehydrogenase	8393381	NP_058702.1	G6pdx	3122660
61	Glucose-6-phosphate dehydrogenase	8393381	NP_058702.1	G6pdx	3122660
173	Glycerol-3-phosphate dehydrogenase 2	6980978	NP_036868.1	Gpd2	3030270
52	P4ha2 protein	34870735	XP_340799.1	P4ha2	
53	P4ha2 protein	34870735	XP_340799.1	P4ha2	
59	P4ha2 protein	34870735	XP_340799.1	P4ha2	
46	Prolyl 4-hydroxylase alpha subunit	51036657	NP_742059.2	P4ha1	
49	Prolyl 4-hydroxylase alpha subunit	51036657	NP_742059.2	P4ha1	
48	Prolyl 4-hydroxylase alpha subunit	51036657	NP_742059.2	P4ha1	
95	Transaldolase	12002054	AAG43169.1	none	11390181

Spot	Proteins that function in cell motility or structure	Gi	Accession	Gene	PMID Reference
142	Tubulin, beta 5	38014544	AAH60540.1	Tubb5	12631274; 3920212
84	Cytoplasmic beta-actin	13592133	NP_112406.1	Actb	9067634; 7835422; 12043959
87	Beta-actin.	55575	CAA24528.1	none	9067634; 7835422; 12043959
101	Gamma-actin	178045	AAA51580.1	ACTG1	9067634; 7835422; 12043959
179	Myosin regulatory light chain 2-A, smooth muscle isoform (Myosin RLC-A)	127170	P13832	Rlc-a	6313687; 7715706
86	put. Beta-actin (aa 27-375).	71620	ATRTC	none	9067634; 12043959; 7835422
89	put. Beta-actin (aa 27-375)	71620	ATRTC	none	9067634; 12043959; 7835422
146	Similar to Cappa1 protein	34859736	XP_227551.2	LOC 310758	4044559
75.2	Similar to tubulin alpha-1 chain – Chinese hamster	27664758	XP_217040.1	LOC 300217	by similarity 3036806
11	Similar to Vinculin (Metavinculin)	27673645	XP_223781.1	LOC 305679	8797806; 6225775; 6417142
12	Similar to Vinculin (Metavinculin)	34868946	XP_223781.2	LOC 305679	8797806; 6225775; 6417142
13	Similar to Vinculin (Metavinculin)	34868946	XP_223781.2	LOC 305679	8797806; 6225775; 6417142
51	Similar to Vinculin (Metavinculin)	27673645	XP_223781.1	LOC 305679	8797806; 6225775; 6417142
43.2	T-plastin	48734834	AAH72523.1	none	9881671

41	T-plastin	2493465	Q63598	PLS3	9881671
76.2	Tubulin alpha	223556	0812252A	none	3036806
77	Vimentin	14389299	NP_112402.1	Vim	1850997; 8373419; 7983050; 8718673; 7832980; 2500966
78	Vimentin	14389299	NP_112402.1	Vim	1850997; 8373419; 7983050; 8718673; 7832980; 2500966
131	Vimentin	38197662	AAH61847.1	none	1850997; 8373419; 7983050; 8718673; 7832980; 2500966
75.1	Vimentin	14389299	NP_112402.1	Vim	1850997; 8373419; 7983050; 8718673; 7832980; 2500966
76.1	Vimentin	14389299	NP_112402.1	Vim	1850997; 8373419; 7983050; 8718673; 7832980; 2500966
Spot	Proteins involved in protein degradation	Gi	Accession	Gene	PMID Reference
133	Cathepsin B	203648	AAA40993.1	None	10876156
138	Huntington interacting protein 2	27695307	XP_214043.1	HIP2	15117950
17	Protease, serine, 15	19173766	NP_596895.1	Prss15	
134	Proteasome (prosome, macropain) subunit, alpha type 3	8394066	NP_058976.1	Psma3	14583091
150	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 6	38014574	AAH60594.1	Serpib6	

154	Similar to serine (or cysteine) proteinase inhibitor, clade B, member 1b; serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 1b	34875374	XP_214459.2	LOC291091	
Spot	Proteins involved in protein synthesis	Gi	Accession	Gene	PMID Reference
97	Acidic ribosomal phosphoprotein PO	57708	CAA33199.1	none	10706406; 8620032
98	Acidic ribosomal protein P0	11693176	NP_071797.1	Arbp	10706406; 8620032
44	Heterogeneous nuclear ribonucleoprotein K	38197650	AAH61867.1	Hnrpk	10329716; 10506147; 11231586; 12052863
80	Heterogeneous nuclear ribonucleoprotein K	38197650	AAH61867.1	Hnrpk	10329716; 10506147; 11231586; 12052863
79	Hnrpk protein	38197650	AAH61867.1	Hnrpk	10329716; 10506147; 11231586; 12052863
107	Translation initiation factor	1240053	CAA58316.1	none	7590282; 7665584; 778232; 8505316; 3112145
Spot	Proteins that act as signalling molecules	Gi	Accession	Gene	PMID Reference
99	Annexin A4	55742832	NP_077069.3	Anxa4	2956093
115	ADP-ribosylation factor 3	18266716	NP_543180.1	Arf3	
96	Annexin 1	6978501	NP_037036.1	Anxa1	1832554
169	Annexin A4	55742832	NP_077069.3	Anxa4	2956093
40	Annexin A6	13994159	NP_077070.1	Anxa6	1420329

42	Annexin A6	13994159	NP_077070.1	Anxa6	1420329
43.3	Annexin A6	13994159	NP_077070.1	Anxa6	1420329
160	Annexin III – rat	71766	LURT3	none	1692066; 2550491
82	cAMP-dependent protein kinase type I-alpha regulatory subunit	6981396	P09456	Prkar1a	10422841
81	Dynactin 2	34865747	XP_216906	Dctn2	12119357; 9126359
147	G-protein beta 2 subunit	41351301	AAH65579.1	none	8809067
132	Lipocortin V	2981437	AAC06290.1	none	10329451
108	Ras-related protein	44890248	AAH66662.1	Rab1	1902553
139	Similar to Ras-related protein Rab-1B	27709432	XP_229035.1	LOC317550	19025532
176	TNF receptor-associated protein 1	55741837	XP_213218	Trap1	8051124
Spot	Other proteins	Gi	Accession	Gene	PMID Reference
114	Dismutase	818029	CAA29121.1	none	
141	Reticulocalbin 1	34856626	XP_342482.1	Rcn1	11886868
174	TUC-4b	21666559	AAM73758.1	none	

**Table 1. Proteins contained in the phosphoprotein fraction**

Each protein is accompanied by its spot number, referred to Figure 14. The Accession number identifies the protein sequence and is often accompanied by the version number, which tracks changes to the sequence itself (Accession.Version); the GI number (genInfo identifier) is the unique integer which identifies a particular sequence and it changes every time the sequence is changed in the database. The name of the correspondent gene is reported when known. The PubMed ID (PMID) references of the papers containing evidence for their phosphorylations are listed.

## **Comparison of IMAC-extracted proteins with the rat genome**

Adsorption of acidic peptides to the affinity matrix has been a limitation in the extraction of phosphopeptides with IMAC (see discussion). It is difficult to prove that none of the extracted proteins result from affinity for the IMAC columns of proteins with a high content of acidic residues. To provide further evidence that the IMAC extraction protocol does not systematically extract such proteins, properties of the extracted proteins with properties of the set of proteins with pI 4-7 in the rat genome were compared (analysis done in collaboration with Dr Larry Brown). In particular, proteins in the rat genome that contain potential acidic binding regions for the IMAC affinity matrix were identified. For this purpose, an acidic pattern was defined as a tetra peptide in which at least 3 of 4 residues are Asp or Glu (Figure 15). An isolated acidic motif was defined as a tetra peptide or penta peptide containing 3 acidic residues in acidic patterns, and which could not be extended under either strict or relaxed conditions (see below). There are three types of sequences that qualify as isolated acidic motifs. A peptide sequence containing sequential, overlapping tetra peptide acidic patterns and which contains at least 4 acidic residues has been defined as an acidic cluster (Figure 15). The acidic clusters are classified as either strict clusters (every adjacent tetra peptide along the sequence contains an acidic pattern) or as relaxed clusters (tetra peptides containing acidic patterns overlap, but are not necessarily adjacent along the sequence). The simplest strict and relaxed acidic clusters are shown in Figure 15. In an individual protein, there may be many isolated acidic motifs, strict clusters and relaxed clusters, as indicated in Figure 15 for the protein complexin 2.

The draft of the rat genome (Jan. 2005) contains a total of 8734 proteins with pI 4-7. To provide a protein set for comparison with the experimentally identified proteins, proteins with more than 1300 residues were excluded since the 2D gels used in the present experiments do

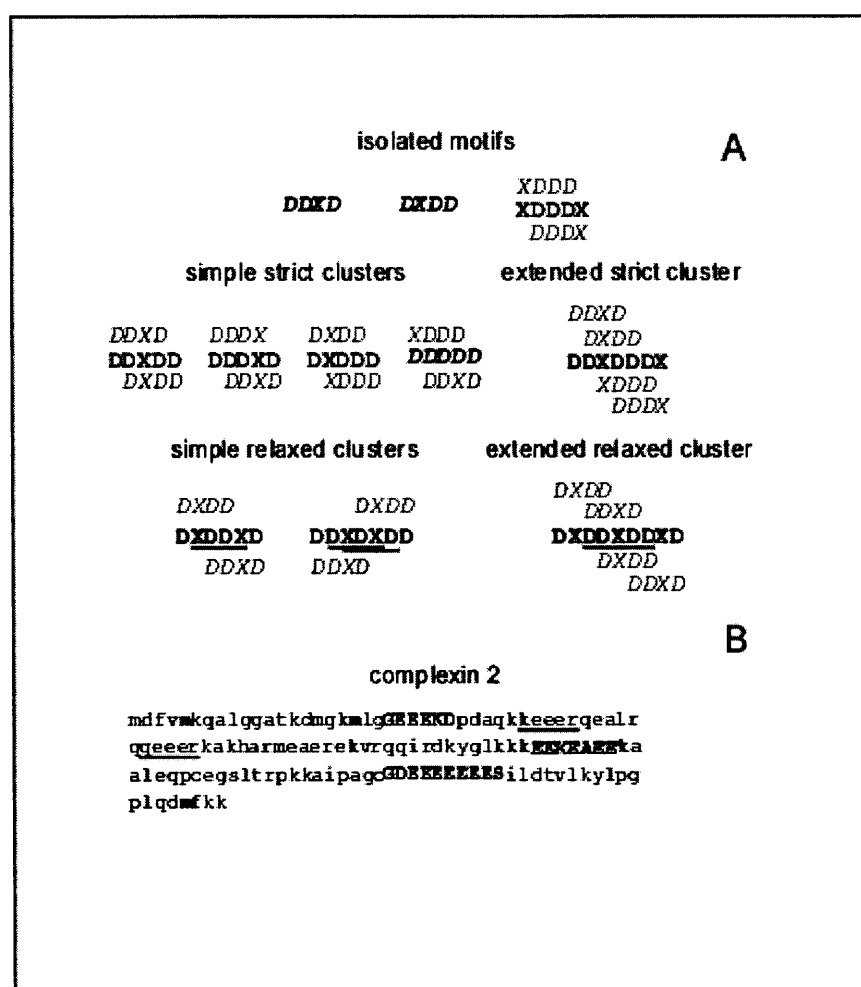
not monitor proteins of very high molecular weight. This left 7971 rat proteins for which all acidic motifs and clusters were identified. A summary of the distribution of motifs and clusters amongst these rat proteins is given in Table 2. Only 15.9% of the rat genome proteins contain neither an isolated motif nor a cluster (Table 2). A further 20.1% contain only a single instance of one or the other of these elements. Although there is an average of about 2 isolated acidic motifs and 1 acidic cluster per rat protein, these are very unevenly distributed in individual proteins with a maximum of 26 isolated acidic motifs (in Gi: 34869308, similar to DnaJ (Hsp40) homologue, subfamily B, member 11) and a maximum of 18 acidic clusters (in Gi: 34858373, similar to ribosome biogenesis protein BMS1 homologue - a handful of proteins with > 1300 residues have still more clusters). Similarly, the size of the acidic clusters varies enormously, with a maximum strict cluster size of 51 residues (in Gi: 19173798, splicing factor YT521-B). Thus, in general isolated acidic motifs and acidic clusters are common in proteins with pI 4-7, with 84.1% of the 7971 proteins containing one or more of these elements.

About half (49.7%) of the rat proteins contain one or more acidic clusters. However, on average those proteins containing acidic clusters have only a slightly greater total content of acidic amino acid residues (Figure 16, A). Acidic clusters are found in proteins with as little as 4.4% total acidic residues and amongst the 166 rat proteins with greater than 21% acidic residues, less than half have acidic clusters. Similarly, proteins containing acidic clusters have on average only slightly lower pI values than the full set of proteins (Figure 16, B) and for both the total proteins and the proteins with acidic clusters, more than half the proteins have pI > 5.5.

Given the ubiquity and diversity of proteins with acidic motifs and clusters, it is no surprise that the proteins experimentally enriched by IMAC contain many such motifs and clusters. Three types of comparisons of the experimentally identified proteins with the rat genome pI 4-



7 protein set were made. First, the distributions of cluster sizes in the two sets of proteins were compared. As shown in Figure 17, A, there is no indication that the IMAC protocol favours extraction of proteins with large acidic clusters. Second, in principle the 3 acidic residues contained in isolated acidic motifs might also constitute sites for binding to the IMAC columns and therefore the distribution of isolated acidic motifs for the two sets of proteins was compared. There is no indication that the IMAC protocol favours extraction of proteins with large numbers of isolated acidic motifs (Figure 17, B). Finally, the distribution of the total number of residues in acidic clusters per protein was compared between the two protein sets. Again there is no indication that high numbers of residues in acidic clusters favours extraction of proteins (Figure 17, C). Indeed, there is a weak indication that large numbers of residues in acidic clusters might be negatively correlated with IMAC extraction. This might reflect negative correlation of phosphorylation with high content of acidic clusters. However, since this is only an average property over all proteins and confirmation would require much more work, this has not been further investigated. In summary, these comparisons provide good evidence that the IMAC extraction protocol is *not* systematically enhancing extraction of proteins with high contents of acidic motifs or clusters and strongly suggest that in proteins where both acidic clusters and phosphorylation exist (Table 2), affinity for the phosphate group is the dominant process under the conditions of this extraction protocol.

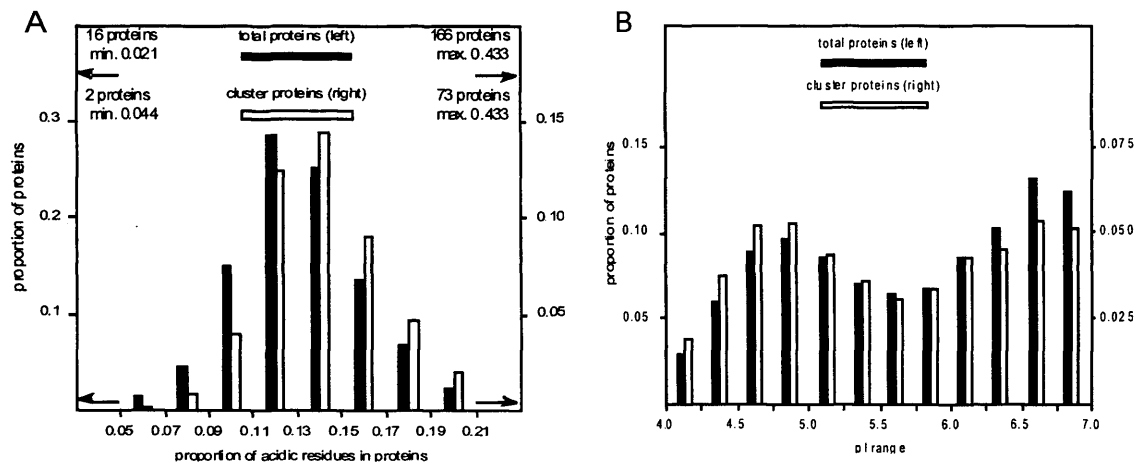


**Figure 15. Schematic description of the definitions used to classify acidic regions in protein sequences**

(A) The primary condition is an acidic pattern in which 3 of 4 residues along the polypeptide sequence are Asp or Glu (patterns in italics under isolated motifs). Strict sequential overlap of the tetra peptide patterns can give rise to strict acidic clusters (every successive tetra peptide along the sequence contains an acidic pattern). In relaxed acidic clusters, the tetrapeptide acidic patterns overlap, but are not necessarily sequential. Relaxed clusters contain tetra peptides that are not acidic patterns and may include sequences that are strict acidic clusters (see the underlined residues). The three types of isolated motifs (bold) are sequences containing 3 acidic residues that can not be extended under either strict or relaxed conditions. (B) Acidic motifs and clusters in the sequence of the protein complexin 2. Isolated acidic motifs are underlined. Strict acidic clusters are shown as bold capitals and relaxed acidic clusters as underlined.

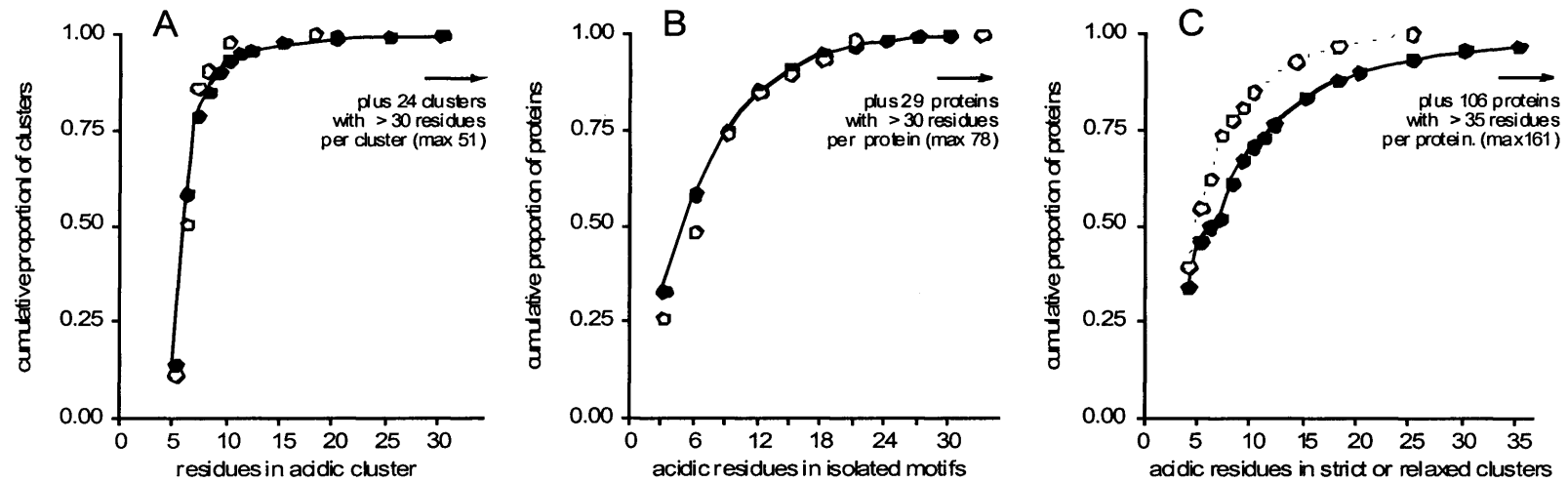
Property	genome set	experimental set
total proteins	7971	62
0 motifs, 0 clusters	1266	5
1 motif, 0 clusters	1180	5
0 motifs, 1 cluster	426	0
> 1 motif, 0 clusters	2747	31
0 motifs, > 1 cluster	652	3
total clusters	8342	42
strict clusters	6799	33
relaxed clusters	1543	9
isolated motifs	16394	169

**Table 2. Distribution of proteins with acidic motifs and clusters**



**Figure 16. Distribution of proteins of the rat genome**

(A) Distribution of proteins over their content of acidic amino acid residues. (B) Distribution of proteins over their pI value.



**Figure 17. Comparison of the properties of isolated acidic motifs and acidic clusters for the experimentally identified proteins (black circles) with proteins in the rat genome (empty circles)**

(A) Proportion of clusters (strict plus relaxed) containing the indicated number of amino acid residues. (B) Proportion of proteins containing the indicated number of amino acid residues in isolated acidic motifs. (C) Proportion of proteins containing the indicated number of amino acid residues in acidic clusters (strict plus relaxed). The number of clusters/proteins in the rat genome which fall outside the plotted regions is indicated.

## DISCUSSION

Effective protocols for isolation of phosphoproteins must satisfy several stringent conditions and an optimal choice of absorbent and elution conditions was probably reached in the described procedure. First, since many lower abundance proteins are regulated by phosphorylation, high rejection of non-selective absorption of high abundance proteins is crucial. Second, if these procedures are to be used in analyses of cellular function under different conditions, extraction of the phosphoproteins must be of high yield and reproducible. The reported experiments provide good evidence for effective extraction of phosphoproteins and utility in signalling study will be further confirmed in the next chapters, in conjunction with kinetic studies of the phosphoproteome response of kidney fibroblasts to stimulation with EGF.

The nature of the chelator, the stationary phase and the buffer systems are all potentially important in obtaining selectivity. For this reason, a discussion of some of the background theory and previous work in this field is probably essential to understand some important considerations in the development of this method. Several support matrices were tested, including more recent innovations such as Toyopearl® AF-Chelate-650M (Tosoh Bioscience) (IDA coupled to synthetic polymer-based particles to give high mechanical strength and chemical resistance) and Fractogel® EMD Chelate (Merck) (IDA coupled to a tentacle gel to give higher functional group density, better steric accessibility and higher protein capacity) (Gaberc-Porekar et al. 2001). No major influences of the stationary phase was noticed, which has also been concluded by other groups (Chaga 2001). Amongst polysaccharide-based matrices, IDA-agarose (Perbio Science), NTA-agarose (Qiagen) and TED-agarose (ActiveMotif) were tested. The chelating ligands tris-carboxymethyl ethylene diamine (TED), nitrilotriacetic acid (NTA) and iminodiacetic acid (IDA) offer a maximum of tri- (IDA), tetra-

(NTA) and penta-dentate (TED) coordination sites to the metal ion. In general, tetra-dentate chelating ligands have higher affinities for metal ions than tri-dentate chelators, but they have lower binding capacity due to the loss of one coordination site on the metal ion. This is even more pronounced in a penta-dentate chelating ligand, where only one coordination site on the metal ion is left for protein binding. No significant phosphoprotein enrichment was found with TED as the chelating ligand. This would be consistent with a proposed model of the binding of phosphate groups to Fe(III) in which two unoccupied Fe(III) coordination sites form a four-membered chelating ring that is necessary for high affinity binding (Muszynska et al. 1986; Muszynska et al. 1992). This can not be the case for a penta-dentate chelating ligand like TED. For extraction of phosphoproteins, both the IDA and NTA groups provided strong enrichment and only small differences in performance between the two were observed.

In contrast, the interaction of carboxyl groups with immobilised Fe(III) is promoted by the presence of several carboxyl groups, which very likely are involved in the formation of a multi-point attachment (Chaga 2001; Muszynska et al. 1986; Muszynska et al. 1992; Porath 1990). The different proposed modes of interaction of phosphate and carboxyl groups with immobilised Fe(III) suggest NTA could have better discrimination against acidic proteins than IDA. In particular, provision of additional coordination sites beyond what is necessary for optimal binding of phosphate groups can only favour proteins with acidic clusters at the expense of phosphoproteins. An additional potential advantage of NTA is that Fe(III) chelate adsorbents can function as pseudo-cation- and pseudo-anion-exchangers under specific low-ionic strength buffer conditions. The interaction between proteins and immobilised Fe(III) can involve both coordination and electrostatic contributions under some conditions. Fe(III) ions bound to chelating ligands form hexahedral coordination sphere complexes in water and, in acidic solution, the Fe(III) complexes have a net charge of 0 (NTA and TED) or +1 (IDA). This means that Fe(III) bound to IDA can assume a pseudo-anion-exchange character with

proteins. Shifting pH in the alkaline direction converts the Fe(III) affinity matrix into a pseudo-cation-exchanger, irrespective of the number of carboxylic groups present in the ligand (Porath 1990; Zachariou et al. 2000). Fe(III) bound to NTA-agarose is not capable of acting as an ion-exchanger under acidic conditions and acidic conditions were therefore chosen in combination with the use of NTA over IDA for investigation of optimal buffer conditions for capture and elution of phosphoproteins.

For purification of phosphopeptides using the IDA-Fe(III) system, it has been demonstrated that lowering the pH value leads to co-elution of fewer acidic peptides with the phosphorylated peptides (Posewitz et al. 1999). Below pH 2.0 the affinity for phosphopeptides is decreased, probably due to protonation of the IDA resin. The critical pH above which an increasing number of non-phosphorylated acidic peptides were retained by the Fe(III) IMAC column seemed to be 3.5. This is one reason why a pH value of 3.4 was chosen for binding. Another reason has to do with the distribution of the isoelectric points of proteins. Proteins are generally poorly soluble near their pI (Arakawa et al. 1985) and proteins from eukaryotes show a bimodal distribution in pI values with maxima around pH 5.5 and 9.5 (Weiller et al. 2004). Less than 1% of proteins have a pI value below pH 4. A binding pH of 3.4 should therefore help to retain as many proteins as possible in a soluble state. Finally, a low binding pH should strongly minimise the binding of proteins rich in surface exposed histidine residues. Under the chosen conditions surface histidines should not be able to undergo any significant  $\pi$ - $\pi$  orbital electron donation via the imidazole moiety as both nitrogens should be protonated. Imidazole itself has a low affinity for hard Lewis metal ions and was therefore incorporated into the washing buffer to quench any residual affinity that histidine residues within individual proteins might have for the immobilised Fe(III).

In conclusion, a new method for enriching phosphoprotein has been optimised, and problems previously reported of protein solubility at low pHs and acidic proteins interference were overcome. Information about the specificity and the recovery of phosphoproteins is also given by comparing the silver stained gels with their autoradiography. Although this procedure does not allow full isolation, but rather enrichment of phosphorylated protein, it is still of high importance in the study of different phospho-isoforms participating in the signalling networks, as I show in the next chapters.

The results presented in this chapter are currently in press (Guerrera et al. 2005b).



# **CHAPTER 4**

## **FISHING FOR KINASE SUBSTRATES IN EGF SIGNALLING PATHWAYS**

### **INTRODUCTION**

In eukaryotes, phosphorylation on serine/threonine is much more frequent than phosphorylation on tyrosine, as mentioned in chapter 1. Phosphorylation on tyrosine has been studied intensively, mainly due to the longstanding availability of specific antibodies against phosphorylated tyrosine. However, efficient antibodies recognising phosphorylated serine and threonine have only become available recently. These include antibodies against phosphoserine/threonine in specific consensus sequences recognised *in vivo* by specific kinases. The optimal amino acid sequences phosphorylated by protein serine/threonine kinases can be experimentally analysed by screening peptide libraries for modified peptides after treatment with a specific kinase. The protein kinase substrate motif can be defined by enrichment and sequencing of phosphorylated peptides. There are many records of using anti phosphoserine/threonine antibodies in chasing and identifying substrates of different kinases (Gronborg et al. 2002; Kohn et al. 2003; Marinissen et al. 2001; Matter et al. 2002). Though there are some inherent problems with the use of antibodies, such as specificity and reproducibility of immuno-detection in western blots, these newly introduced antibodies were used in the present study.

Antibodies against phosphorylated substrates of MAPK, PKC and PKA were used for a partial screen of the fibroblast proteome. In detail, the antibodies mainly used and adopted for the following experiments were: (i) against MAPK/CDK substrates, which recognise phosphorylated threonine when followed by Pro (TP); (ii) against PKC substrates, which

recognise phosphorylated serine surrounded by Arg or Lys in the -2 and +2 position, and a hydrophobic residue at + 1 position ([RK]XS[Hyd][RK]); (iii) against PKA substrates, which recognise phosphorylated Ser/Thr with Arg in -3 position (RXXS/ RXXT). For simplicity they are referred to (i) as TP, (ii) as RXS and (iii) as RXXS/T.

All these kinases are known to be involved in the EGF signal cascade (see chapter 1). In an attempt to investigate some of the downstream players in different activated pathways, the immuno response of enriched phosphoproteomes from NRK49F cells stimulated with EGF was compared with the corresponding phosphoproteome from unstimulated cells. This investigation was complemented by densitometric analysis for an indication of the phosphorylation kinetics of the proteins identified.

Three main factors suggest that this analysis must be considered as a proof-of-concept study. First, intrinsic limitations of the techniques and materials used in this approach exclude the possibility of obtaining a wide and exhaustive screening of kinase substrates. Second, the antibodies used cover only a subset of the potential substrates for the kinases selected. Third, a scan of the SwissProt database for the consensus sequence [RK]XS[Hyd][RK] using Patterinprot software revealed that out of the 193 proteins containing this consensus sequence, only 59% would fall in the pI and MW range used in this work. These factors limit the outcome of this screening to be less than exhaustive. Some practical limitations that have been found during this analysis are discussed in details at the end of this chapter.

The EGF-induced phosphorylation changes of the subset of proteins identified as potential MAPK, PKC and PKA substrates were analysed in a semi-quantitative way. These proteins were identified using MALDI-TOF MS. Along with the identification, MALDI-TOF spectra often provide information about the coverage, which is the percentage of primary structure matched by the tryptic peptides experimentally obtained. The distribution of these peptides

along the sequence can also be indicative of possible splice variants. Finally, leads were obtained for potentially phosphorylated peptides, which correspond to peaks associated with a tryptic peptide of the protein plus 80Da or multiples of 80Da (Guerrera et al. 2005a). However, observations on phosphorylation sites were not exhaustive because phosphopeptides are more reluctant to ionise in MALDI-TOF MS. An additional limitation of the overall analysis is that only proteins from the control could be analysed. For these reasons the focus in this chapter is on substrate identification and not on potential phosphorylation sites.

The following chapter describes a more sophisticated mass spectrometric analysis of some proteins identified here. Application of this analysis yielded further details of peptide phosphorylation site and quantitative information on the kinetics that characterise these regulatory events.

## RESULTS

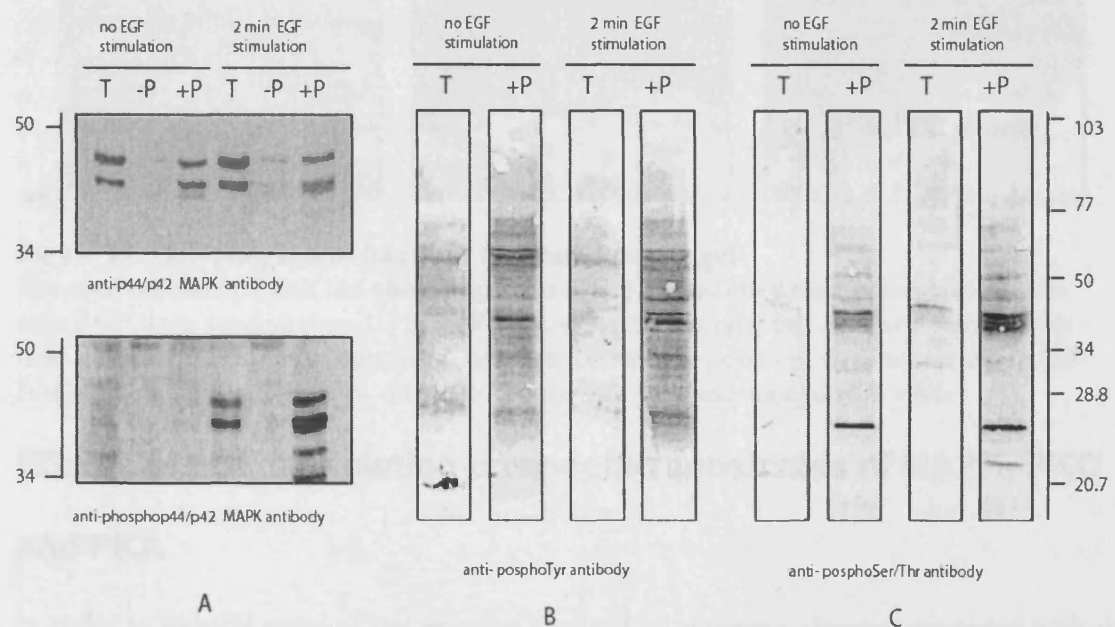
### Effects of EGF stimulation in NRK49F cells

Phosphoproteins were isolated from normal rat kidney fibroblast cells (NRK49F) as described in chapter 3 after stimulation with EGF for 0, 2 and 5 minutes.

Selected signalling proteins specifically phosphorylated in response to EGF were followed in order to verify the efficiency of the stimulation. Equal amounts of phosphoproteins and total proteins were separated by 1D-SDS PAGE and then electroblotted onto a membrane. Proteins blotted were probed with antibodies specific for p44 and p42 mitogen-activated protein kinase (MAPK) proteins and with antibodies specific for a phosphorylated form of the same proteins. P44/p42 are activated through a double phosphorylation, and the activation by phosphorylation, marker of transmission of the signal, was found to be maximal at 2 minutes (data not shown). Phosphorylated p44/p42 was fully recovered during the purification, while total p44/p42 was present in the same amount in both the whole cell lysate and in the phosphoprotein fraction (Figure 18, A). Similar results were obtained with phospho Map38 (data not shown). Furthermore, the phosphorylated p44/p42 signal after stimulation is much stronger in the phosphoprotein fraction than in the total protein, whilst the signal in the unstimulated cell was not improved, confirming the efficiency of the phosphoprotein enrichment. Total MAPK (modified and unmodified) was followed as well to demonstrate that the same amount of protein has been loaded. Interestingly no endogenous MAPK is found in the flow through, suggesting that the protein is lost in the washing steps performed before the elution.

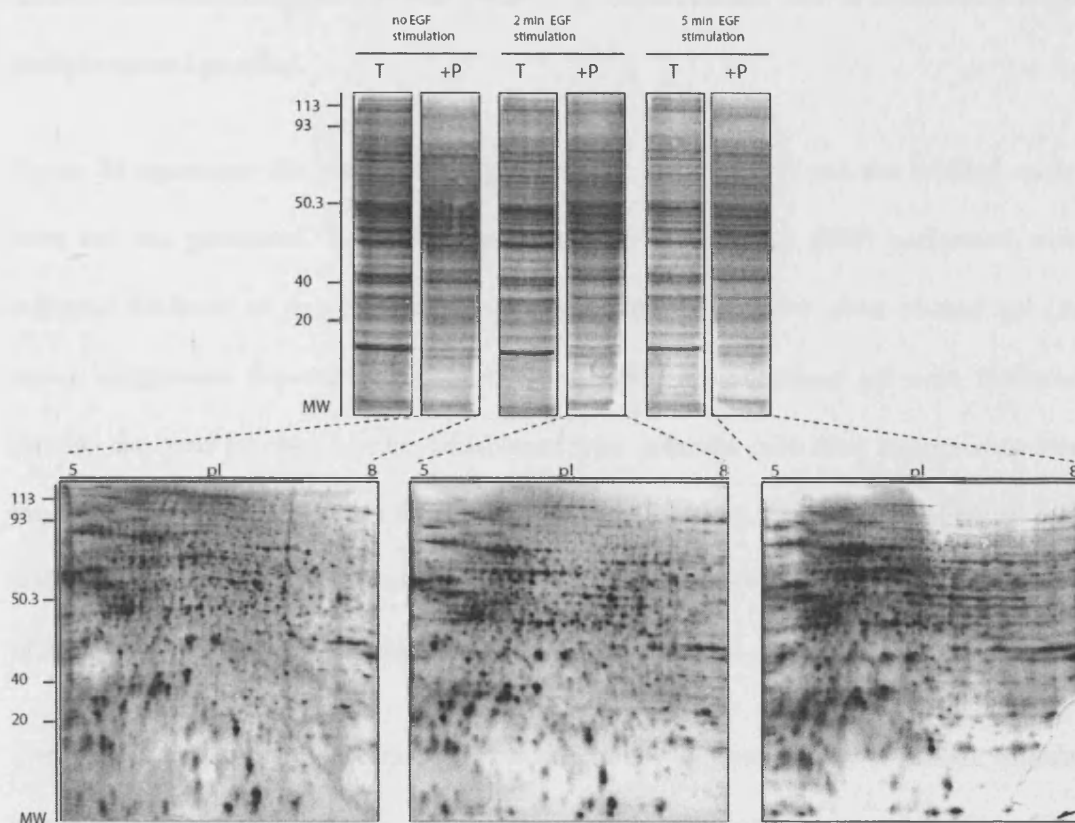
Considerably stronger signals in the phosphoprotein fraction compared to the total protein fraction were observed on blots probed with anti-phospho-tyrosine antibodies and anti-phospho-(Ser/Thr) PKA substrate (Figure 18, B-C).

When total protein and phosphoprotein fractions were stained with comassie, different patterns were obtained on 1D gel. However, pattern changes of the phosphoprotein fraction at different times could hardly be observed, both on 1D and 2D PAGE separation (Figure 19). For this reason, a different experimental route was employed, namely a combination of western blotting and phospho motif specific antibodies. The yield of the phosphoprotein enrichment was 12% and ca. 800 spots were detected by the software in each 2D gel shown here.



**Figure 18. Analysis of enrichment of phosphoproteins after IMAC columns**

Quiescent NRK49F cells were treated with or without EGF for the indicated time periods and protein extracts subjected to enrichments using IMAC columns. Same amounts of total cellular extracts (T), IMAC column flow through (-P) and IMAC column eluate (+P) were separated on SDS-PAGE and subjected to immunoblot analysis. (A) Blot was probed with anti-phospho-p44/42 MAPK (lower panel) and reprobbed with anti-p44/42 MAPK (upper panel) antibodies. (B) Blot was probed with anti-phospho tyrosine antibody. (C) Blot was probed with anti-phospho threonine/serine antibody.



**Figure 19. Phosphoprotein fractions separated on 2D gel**

The total proteins (T) and the phosphoproteins (+P) at described time points stimulation with EGF were separated on 1D SDS-PAGE using 13% acrylamide gels, and stained with blue comassie. Phosphoprotein fractions from the same experiment were separated on 2D-PAGE using 7cm pI 5-8 strips on 11.5% acrylamide gels, and stained with silver.

### **Effects of EGF stimulation on specific substrates of MAPK, PKC and PKA**

In order to identify some of the proteins involved in proteome changes associated with the stimulation, phosphoproteins were resolved on 2D-SDS PAGE and western blotted using antibodies against phosphorylated protein kinase C (PKC), protein kinase A (PKA) and mitogen activating protein kinase (MAPK) substrates.

Three different sets of immunoblots of the NRK49F phosphoproteome at 0, 2 and 5 minutes were obtained. As the same amount of phosphoproteins was loaded at each time point,

different intensities suggest different grade of phosphorylation and/or number of copies of phosphorylated proteins.

Figure 20 represents the silver stained gel used for identification and the labelled spots that were cut and processed. Each panel presents the western blot (WB) performed with the indicated antibody of the corresponding area on the preparative silver stained gel (dotted areas). Alignments between blots and the preparative silver stained gel were facilitated by staining the total proteins on the membranes with colloidal gold after immunodetection. In this way the protein patterns on the membranes were matched with the ones obtained by silver staining (using Photoshop software). As a result, the proteins of interest visualised by western blotting could be properly allocated on the preparative gel.

The gel containing the phosphoproteins obtained from the control cells (0 minute stimulation) was used for the analysis and for the identification of the proteins targeted by the antibodies. Significant protein spots were excised from the gel, washed and digested *in situ* with trypsin. The resulting peptides were analysed using MALDI-TOF MS and the proteins identified are listed in Table 3. Despite the fact that a broad range of high and low abundance proteins was observed when the membranes were stained with colloidal gold (data not shown), the protein pattern in western blots matched mostly high abundance proteins. This suggests that high phosphorylation level on these spots might obscure the signal of very low abundant proteins. This issue was not followed further because the detection of low abundance substrates by WB probably could have not been complemented with satisfactory identification.

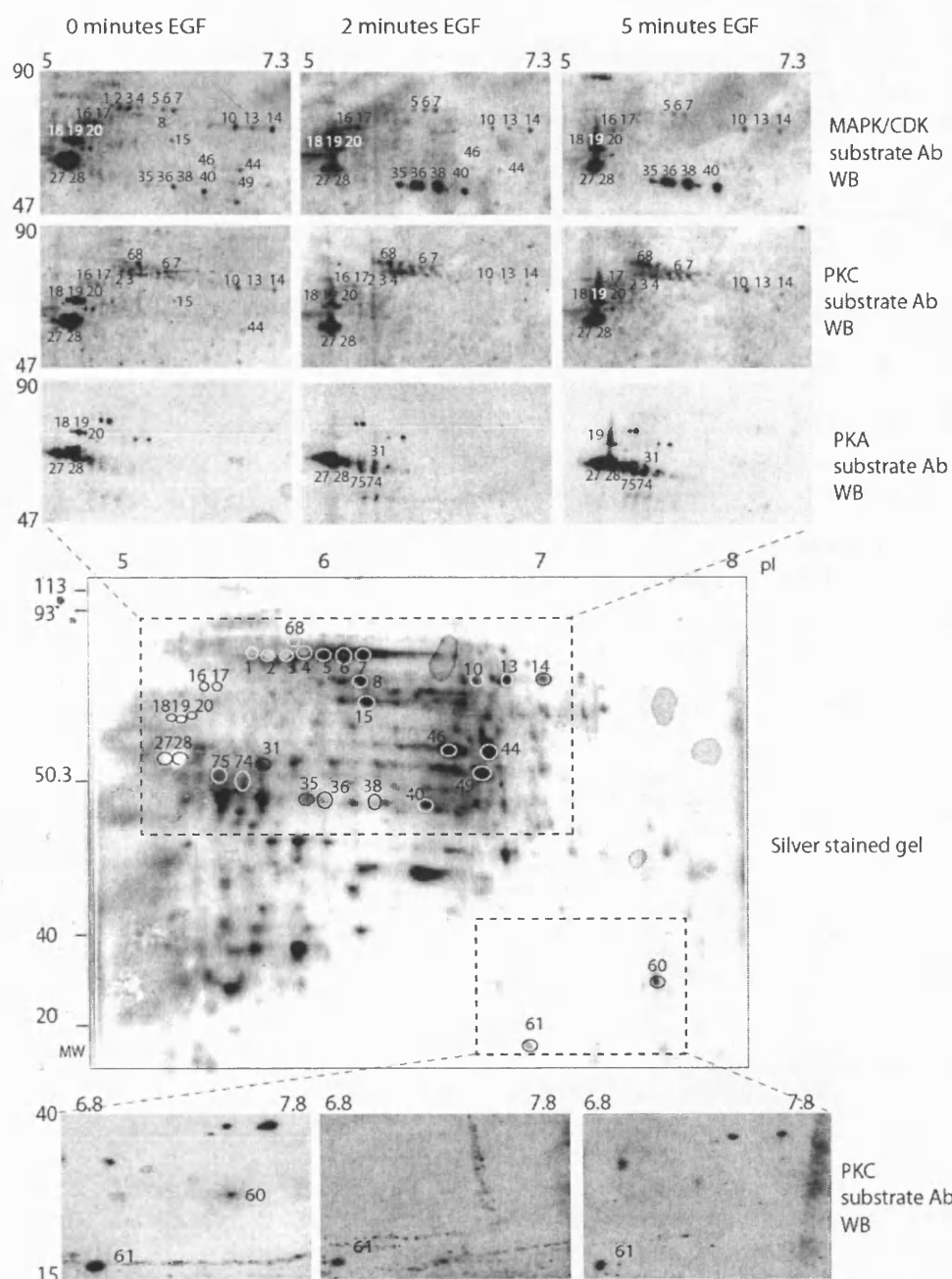
In summary 21 spots out of 33 analysed could be identified; these spots contained a total of 9 different proteins. The discrepancy is due to the presence of different isoforms of the same protein (the same protein found in more than one spot). As commented before, the variation in isoelectric point is usually associated with different degrees of post translational

modifications (PTMs), which may or may not cause a significant shift in molecular weight. In this experiment, as all isoforms were detected with antibodies against phosphorylation, at least one of the PTM in question must be phosphorylation. In the opposite case, some spots contained more than one protein, therefore mixtures of more than 2 proteins have not been considered because no reliable connection with the phosphorylation kinetics could be made. 90% of the identifications are characterised by a minimum of 15% coverage. In few cases the coverage was lower, but the protein identifications were still considered valid by the software performing the search. This is probably due to the low number of peptides submitted to the search and to the high accuracy of the match.

Some proteins could not be identified due to their very low abundance in the cell, (therefore low intensity on the gel). Amongst these, spots 35, 36 and 38 showed an interesting behaviour as MAPK/CDK substrates (Figure 20). Spot 16 and 17, which are substrates of MAPK/CDK and PKC could not be identified either, as they were not even detectable on the silver stained gel. Interestingly, there are two spots in the PKA substrate blots in the area where 16 and 17 are found in the other western blots, but they are strongly shifted toward the alkaline region and have lower molecular weight. These could possibly be the same proteins, but no speculation could be made as they have not been identified. For this reason these spots are not labelled as 16 and 17.

shift 0.04 unit





**Figure 20. Mapping of kinase substrates on 2D gel**

The central panel represents the silver stained gel of the phosphoprotein fraction of unstimulated cells. This gel was used for the cutting and the analysis of the spots of interest. The phosphoprotein fractions obtained from cells EGF stimulated for 0, 2 and 5 minutes were separated on 2D PAGE using a 13cm pI 4-7 strip, on a 13% acrylamide gel. The phosphoproteins were transferred to membranes and probed with the indicated phospho substrates antibody. The dotted areas on the gel correspond to the zooms on the relevant areas of the different western blots.

s	gene	Accession number	Gi	Protein	Cov.	PM	Ppm
1	Hspa8	NP_077327	13242237	Heat shock protein 8	32%	19	32
2	Hspa8	NP_077327	13242237	Heat shock protein 8	39%	37	36
3	Hspa8	NP_077327	13242237	Heat shock protein 8	19%	9	50
4	Hspa8	NP_077327	13242237	Heat shock protein 8	22%	11	15
7	NM_130894	NP_570964	29501815	Hypertension-related protein; mitochondrial transmembrane GTPase	12%	7	71
8	Hspa8	NP_077327	13242237	Heat shock protein 8	36%	21	26
13	Stip1	NP_620266	20302113	Stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organising protein)	21%	15	94
14	Stip1	NP_620266	20302113	Stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organising protein)	17%	9	83
15	Pdia3	P11598	1352384	Protein disulfide-isomerase A3 precursor	22%	11	25
18	Vim	NP_112402	14389299	Vimentin	49%	21	14
	Tuba1	NP_071634	11560133	similar to Tubulin alpha-1 chain	25%	10	23
19	Vim	NP_112402	14389299	Vimentin	52%	23	19
	Tuba1	NP_071634	11560133	similar to Tubulin alpha-1 chain	26%	10	31
20	Vim	NP_112402	14389299	Vimentin	50%	23	55
	Tuba1	NP_071634	11560133	similar to Tubulin alpha-1 chain	23%	9	63
27	Vim	NP_112402	14389299	Vimentin	60%	34	27
	P4hb	P04785	129731	Protein disulfide isomerase precursor (PDI)	28%	13	35
28	Vim	NP_112402	14389299	Vimentin	58%	32	29
	P4hb	P04785	129731	Protein disulfide isomerase precursor (PDI)	21%	10	38
31	Hspa5	P06761	121574	78 kDa glucose-regulated protein precursor (GRP 78) (BiP)	15%	8	37
40	Hspa8	NP_077327	13242237	Heat shock protein 8	10%	5	15
46	Eno1	NP_036686.1	6978809	Enolase-1, alpha	34%	13	55
49	Eno1	NP_036686.1	6978809	Enolase-1, alpha	40%	17	48
60		AAB19888	8248819	Phosphoglycerate mutase type B subunit	15%	5	23
61	Gstp1	P04906	121749	Glutathione S-transferase P (Chain 7)	34%	5	37
68	Hspa8	NP_077327	13242237	Heat shock protein 8	21%	9	177

**Table 3. List of proteins identified**

S is the number of the spot; the Accession number identifies the protein sequence and is often accompanied by the version number, which tracks changes to the sequence itself (Accession.Version); the GI number (genInfo identifier) is the unique integer which identifies a particular sequence and it changes every time the sequence is changed in the database; Cov is the % of the protein found as tryptic peptides; PM is the number of peptides matched, ppm is the maximum error in part per million.

## Phosphorylation kinetics

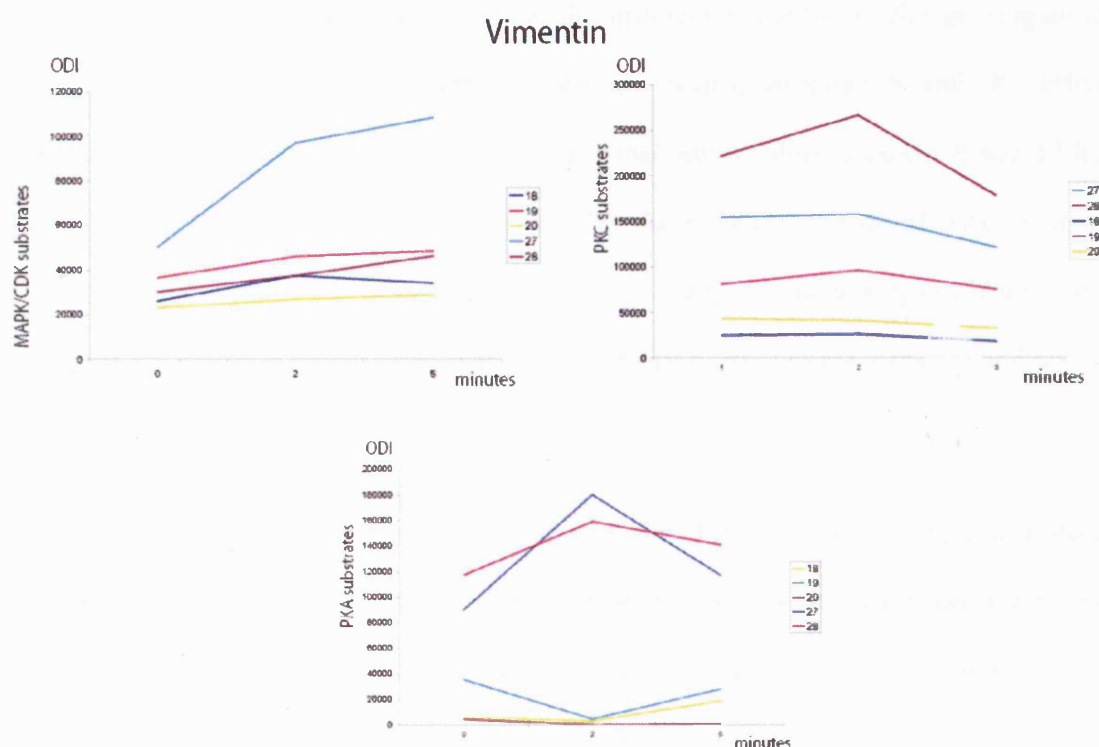
When same amounts of proteins are used, the intensity of the signal obtained by WB using antibodies against phosphorylated residues is very likely correlated with the degree of protein phosphorylation. In order to describe the EGF related phospho-kinetics of the proteins identified, changes in intensities of the proteins spots from the exposed films were quantified using ImageQuant software. Each spot was calibrated against a large area in the background of the blot, and equal areas were taken when spots were compared from one blot to the other. Comparisons were only performed for spots belonging to the same set of western blots, immunostained with the same antibody and developed for the same length of time. Finally, the intensities of the spots were normalised to the protein loading, by comparing the intensities of the spots from the corresponding membranes stained with colloidal gold.

The proteins analysed were grouped according to their role in the cell, and a detailed analysis of their phosphorylation kinetics is reported below.

### *Structural proteins*

#### *Vimentin isoforms*

All identified vimentin isoforms were recognised by all antibodies used, and a similar phosphorylation kinetic characterised the spots from the same set of blot (Figure 21). Spots 18, 19 and 20 contained Vimentin together with alpha tubulin, and spots 27 and 28 contained vimentin together with Protein Disulfide Isomerase (PDI) precursor. Although the mixture was dominated by vimentin (according to the respective coverage of the two proteins in the same spot) these kinetics could not be considered a reliable quantification of the phospho-kinetics of vimentin.



**Figure 21. Phospho-kinetics of Vimentin**

Each curve represents one protein isoform of Vimentin, called with the name of the corresponding spot on the western blots in Figure 20. The diagrams report the ODI (optical density intensity) of each isoform, which corresponds to their behaviour as substrates of the indicated kinases (y axis) versus the minutes of stimulation with EGF (x axis).

### *Proteins involved in chaperone regulated redox network*

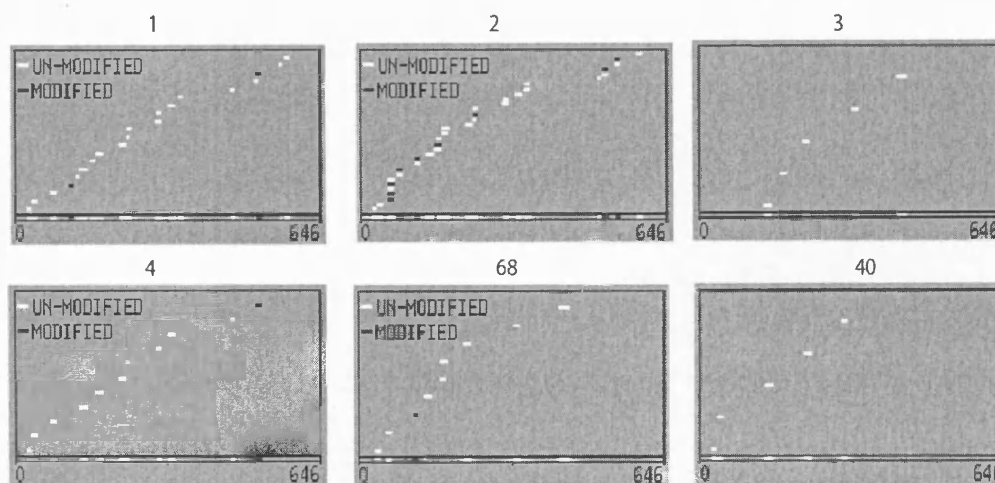
#### *Heat Shock Protein 8 isoforms*

For Heat Shock Protein 8 (Hsp8) different isoforms were observed, and they behaved as distinct substrates of MAPK/CDK and PKC. Indeed, whilst the isoforms contained in spots 2, 3 and 4 were recognised by both antibodies, spots 1, 8 and 40 were specifically recognised by the MAPK/CDK substrate antibody and spot 68 was recognised only by the PKC substrate antibody. More proteins were present in this area of the anti PKC substrate blots (Figure 20) but only spots visible on the silver stained gel and successfully identified are reported in these graphics.

As readily noticeable from the positions of the different isoforms on the gel (Figure 20) although they all had slightly different isoelectric points, isoforms 8 and 40 differed substantially in molecular weight (MW) from the theoretical values. Isoform 8 was 10 kDa lighter whilst isoform 40 was smaller by 30 kDa. Interestingly, the distribution of tryptic peptides along the amino acid sequence of the protein changed considerably only for isoforms 40 (Figure 22): the peptides matched only the N-terminal portion of the sequence and this was in agreement with the difference in MW observed on the gel.

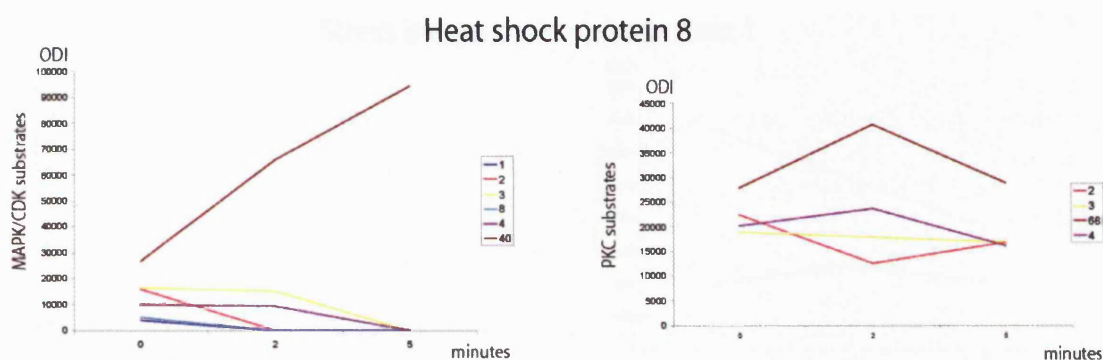
Kinetics of differential phosphorylation are reported in Figure 23; interestingly all isoforms underwent dephosphorylation on TP with different rates, except 40, which became strongly phosphorylated. Phosphorylation on potential sites recognised by anti phospho PKC substrates did not seem to be overall influenced by EGF treatment.

Quantitation of differential phosphorylation on different sites of Hsp8 is described in the next chapter.



**Figure 22. Distribution of tryptic peptides in different isoforms of Hsp8**

The x axes represent the aminoacidic sequence of Hsp8; the segments in the graphic represent the tryptic peptides matched to the sequence; modified stands for peptides containing oxidised methionine and/or cysteine alkylated.

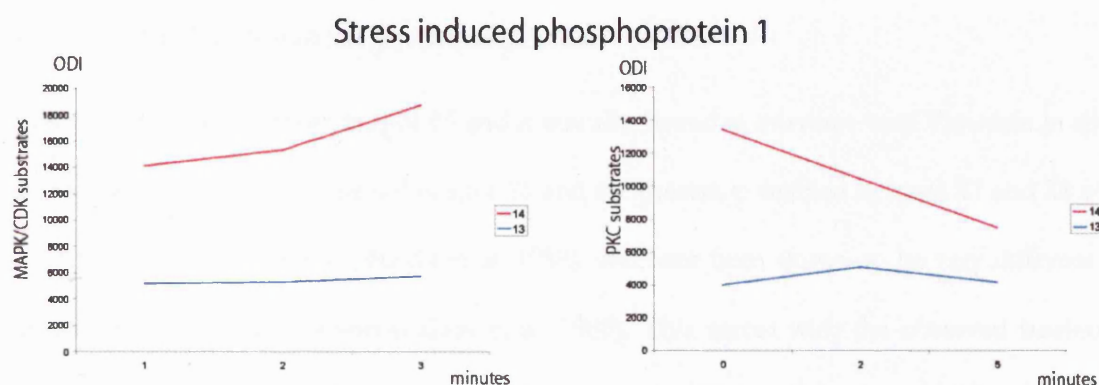


**Figure 23. Phospho-kinetics of Hsp8**

Each curve represents one protein isoform of Hsp8, called with the name of the corresponding spot on the western blots in Figure 20. The diagrams report the ODI (optical density intensity) of each isoform, which corresponds to their behaviour as substrates of the indicated kinases (y axis) versus the minutes of stimulation with EGF (x axis).

### *Stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organising protein) isoforms*

Hsp70/Hsp90-organising protein (corresponding to human Hop) was found in spot 13 and 14. The predicted molecular weight and pI for this protein are 63 kDa and pI 6.4. Spots 13 and 14 showed the same molecular weight (around 65kDa), but a difference in 0.2 pI units (pI 6.9 and pI 7.1 respectively). The isoform contained in spot 13 did not change its phosphorylation state, whilst phospho-isoform 14 seemed to increase phosphorylation on TP, but to decrease it on RXS (Figure 24). The sequence contains Thr332 and Thr398 which are followed by proline (TP).

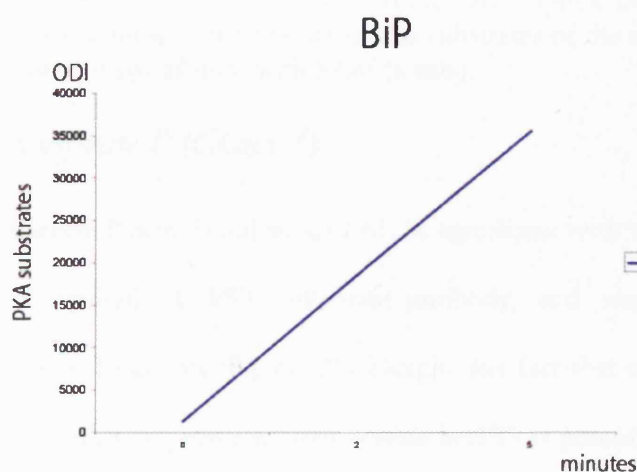


**Figure 24. Phospho-kinetics of Stress induced phosphoprotein 1**

Each curve represents one protein isoform of stress induced phosphoprotein 1, called with the name of the corresponding spot on the western blots in Figure 20. The diagrams report the ODI (optical density intensity) of each isoform, which corresponds to their behaviour as substrates of the indicated kinases (y axis) versus the minutes of stimulation with EGF (x axis).

### *78 kDa glucose-regulated protein precursor (GRP 78) (BiP)*

BiP is also known as Heat shock protein 5; it was found in spot 31, but presented quite a lower molecular weight (55kDa) than predicted (72kDa) and a higher pI (pI 5.7 rather than pI 5.1). This protein underwent strong and clear EGF-dependent phosphorylation on RXXS/T (Figure 25). Ser300 could be a potential site because it lies within a typical PKA consensus motif.

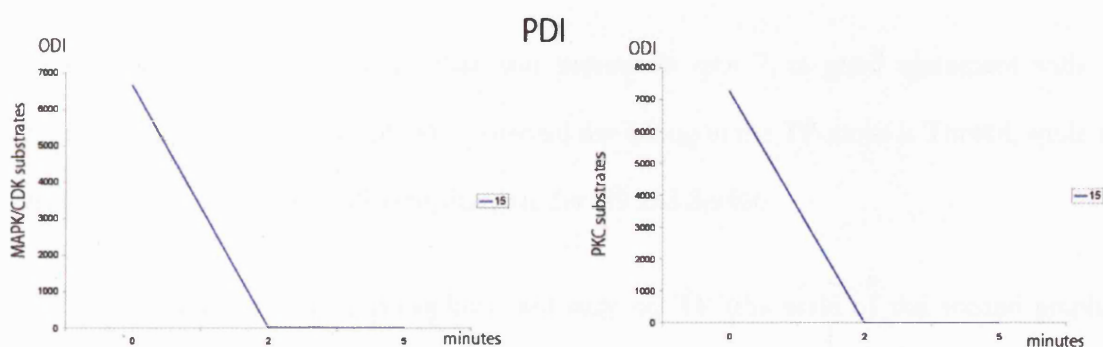


**Figure 25. Phospho-kinetics of BiP**

This curve represents BiP, called spot 31 according to the western blots in Figure 20. The diagram reports its ODI (optical density intensity), which corresponds to the behaviour as substrates of PKA (y axis) versus the minutes of stimulation with EGF (x axis).

### *Protein disulfide-isomerase precursor proteins*

PDI precursor was present in spot 15 and it was also found in mixtures with Vimentin in spots 27 and 28. The protein contained in spot 15 and the protein contained in spots 27 and 28 were first thought to be identical (Boado et al. 1988), and later been shown to be very different on primary structure level (Schoenmakers et al. 1989). This agrees with the observed isoelectric points but not for molecular weights of these proteins. The isoform in spot 15 presented a remarkably higher MW (ca 15 kDa heavier than predicted), and it was quickly dephosphorylated upon stimulation with EGF (Figure 26). PDI contains a TP motif at Thr462, and Ser150 in [RK]XS[Hyd][RK] (but with K in -3 rather than in -2).



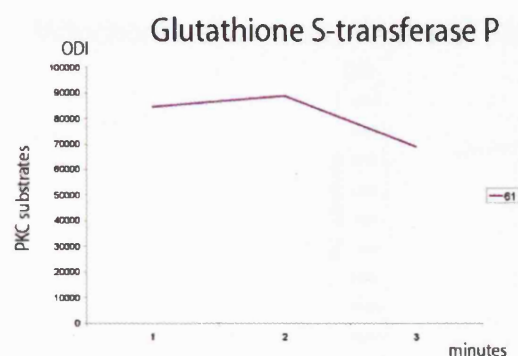
**Figure 26. Phospho-kinetics of PDI**

Both curves represent the protein PDI, called with the name of the corresponding spot on the western blots in Figure 20. The diagrams report the ODI (optical density intensity) of this protein, which corresponds to their behaviour as substrates of the indicated kinases (y axis) versus the minutes of stimulation with EGF (x axis).

### *Glutathione S-transferase P (Chain 7)*

Glutathione S-transferase P was found in spot 61, in agreement with the theoretical MW and pI. It was only recognised by PKC substrate antibody, and seemed to undergo light dephosphorylation after 2 minutes (Figure 27). Despite the fact that no phosphopeptide was observed in the spectrum, a sequence analysis reveals Ser185 as potential phosphorylation site for PKC.





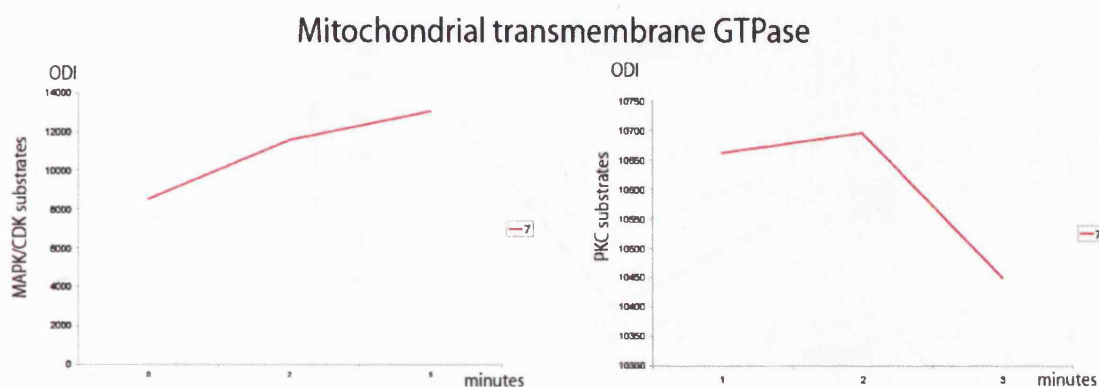
### Figure 27. Phospho-kinetics of Glutathione S-transferase P

This curve represents Glutathione S-transferase P, corresponding to spot 61 in the western blots in Figure 20. The diagram reports its ODI (optical density intensity), which corresponds to the behaviour as substrates of PKC (y axis) versus the minutes of stimulation with EGF (x axis).

### *Mitochondrial transmembrane GTPase*

Mitochondrial transmembrane GTPase was present in spot 7, in good agreement with the theoretical values of MW and pI. The potential site falling in the TP motif is Thr484, while the candidate residues for [RK]XS[Hyd][RK] are Ser149 and Ser466.

This protein appeared to be phosphorylated only on TP (the scale of the second graph is strongly expanded) (Figure 28).



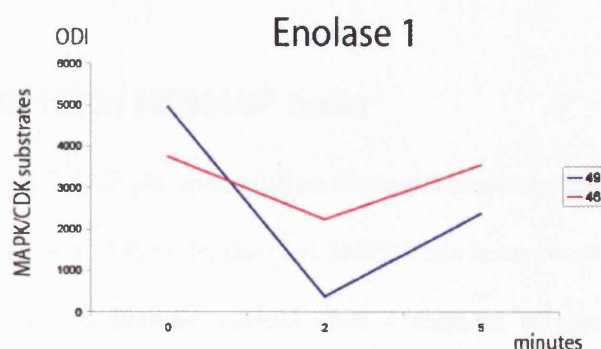
**Figure 28. Phospho-kinetics of mitochondrial transmembrane GTPase**

Both curves represent the protein mitochondrial transmembrane GTPase, called with the name of the corresponding spot on the western blots in Figure 20. The diagrams report the ODI (optical density intensity) of this protein, which corresponds to their behaviour as substrates of the indicated kinases (y axis) versus the minutes of stimulation with EGF (x axis).

## Metabolic enzymes

### *Enolase-1 isoforms*

Enolase-1 showed a similar conduct for both isoforms contained in spots 46 and 49 as substrates of MAPK/CDK. Interestingly, a spot adjacent to spot 46 contained Enolase-1 (in mixture) which is not recognised by the antibodies used, but that is likely to be phosphorylated because enriched with IMAC columns. The coverage obtained for the different isoforms was high, and the distribution on the sequence very similar. This agreed with the observation that the three isoforms migrate at similar molecular weight, although this was 10kDa higher than predicted; as for the different isoelectropoint, the isoforms showed similar general behaviour (Figure 29).

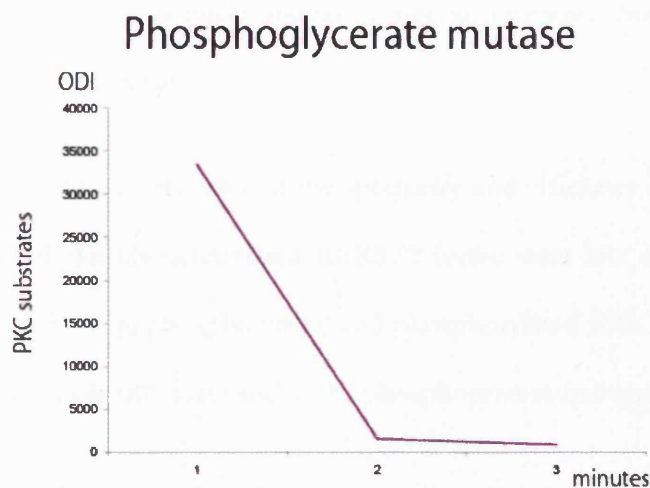


**Figure 29. Phospho-kinetics of Enolase-1**

Each curve represents one protein isoform of Enolase-1, called with the name of the corresponding spot on the western blots in Figure 20. The diagrams report the ODI (optical density intensity) of each isoform, which corresponds to their behaviour as substrates of MAPK/CDK (y axis) versus the minutes of stimulation with EGF (x axis).

### *Phosphoglycerate mutase type B*

Phosphoglycerate mutase type B was found only in spot 60; its observed isoelectric point was 0.6 pI units lower than the theoretical, while its MW was in good agreement. This protein is recognised as being phosphorylated on [RK]XS[Hyd][RK] at time 0, which could be Ser118 (although with only 76% similarity with the strict recognised sequence) (Figure 30).



**Figure 30. Phospho-kinetics of Phosphoglycerate mutase**

This curve represents Phosphoglycerate mutase, corresponding to spot 61 in the western blots in Figure 20. The diagram reports its ODI (optical density intensity), which corresponds to the behaviour as substrates of PKC (y axis) versus the minutes of stimulation with EGF (x axis).

## **DISCUSSION**

### **EGF activates ERK1/2 in NRK49F cells**

In fibroblasts, EGF induces ERK2 phosphorylation through alternative pathways in a cell type specific manner (Burgering et al. 1993). In the past MAPK has been shown to be activated in this cell line, with maximum activation reached after 4 minutes of stimulation with EGF (Lahaye et al. 1998).

Similar amounts of total cellular proteins (proteome) and total phosphoproteins (phosphoproteome) were loaded on 1D-gels and changes of ERK1/2 and p38MAPK were followed in both protein fractions. IMAC proved to enrich the phosphoprotein fraction; after 2 minutes of stimulation with EGF the phospho Erk1/2 signals were much stronger in the phosphoprotein fraction than in the total protein fraction, whilst the signal from the unstimulated cells was not enhanced (Figure 18, A). The results obtained suggest that EGF activates ERK1/2 and p38MAPK in NRK49F cells within the first 2 minutes of cell stimulation (Figure 18, A). This effect decreased later, at 5 minutes, but was still present even after 10 minutes (data not shown).

Some conclusions could be drawn about the specificity and efficiency of the phosphoprotein enrichment. None of the phosphorylated ERK1/2 forms were lost during the purification, while the mixture of both unphosphorylated and phosphorylated ERK1/2 was present in the same amount in the whole cell lysate and in the phosphoprotein fraction.

In summary, the results obtained so far in this study show that ERK1/2 are activated as a response of the NRK49F cells to EGF stimulation under the conditions used and that the IMAC based methodology developed is appropriate for further studies of signalling related phosphorylation events.

## **Phosphoproteome changes upon EGF stimulation**

No visible changes could be seen in the proteome and in the phosphoproteome of NRK49F after stimulation with EGF when proteins were separated on 1D SDS-PAGE (Figure 19). The NRK49F phosphoproteins after 0, 2 and 5 minutes of stimulation with EGF were therefore separated on 2D gels, but in spite of the high resolution of the gels obtained the analyses and comparisons of the gels appeared too laborious and complex with the software available (Figure 19). This seems to be recognised as a common problem, and in spite of the advances in gel electrophoresis and software performance it is quite unlikely that 2D gels can be exploitable to compare and analyse complex samples with high accuracy (Mann et al. 2001).

However, classical proteomics approaches have historically relied on 2D PAGE separation, although this approach encounters the above mentioned limitation as well as others, such as poor reproducibility and failure to resolve proteins above 200 kDa, proteins with very acidic and alkaline pI and membrane proteins. In spite of these restrictions, 2D-PAGE is still the most powerful method for separating proteins with high resolution, and was chosen in this study in combination with a higher level of partition.

## **Kinase substrates immunodetection and identification**

The largest group of enzymes that catalyse regulatory posttranslational modifications is the family of protein kinases. These enzymes transfer phosphate groups from ATP to specific hydroxyl groups on substrate proteins.

To obtain additional information concerning the EGF-dependent phosphorylation on some proteins, the phosphoproteome was screened for phosphorylated residues contained in consensus sequences known to be recognised by certain families of kinases. The choice of antibodies was based on previous work describing the role of EGF in activating the MAPK

pathway (see chapter 1). Phosphospecific antibodies have recently been introduced as tool in a new approach to protein kinase substrate identification (described in the introduction to this chapter). In this study these antibodies have been used for a general screen of phosphoproteomes. Subclasses of serine/threonine phosphorylated proteins that could potentially behave as protein kinase substrates have been identified. After stimulating cells with EGF for 0, 2 and 5 minutes, phosphorylation changes in multiple proteins were followed using antibodies specific for phosphorylated threonine followed by proline (TP) (MAPK/CDK phosphorylated substrates), phosphorylated serine in RXS consensus (PKC phosphorylated substrates), and both serine and threonine in RXXS/T consensus (PKA phosphorylated substrates). However, it must be added that, even assuming the antibody specificity is very high, this approach would not capture atypically phosphorylated proteins, proteins out of the pI/MW range used and low abundant proteins.

Phosphoprotein fractions for cells treated with EGF were separated on 2D gels and specifically immunostained (Figure 20). Blot images showed significant changes in the pattern of protein spots after stimulation with EGF for 2 and 5 minutes. This confirms that EGF causes a boost in phosphorylation events in the cell. Upon stimulation, some proteins become more phosphorylated, others become dephosphorylated while some remain unchanged. This is consistent with previous work showing that EGF activates phosphatases as well as kinases (Cunnick et al. 2002; Ren et al. 2003). Some of the proteins appeared in more than one blot, justifiable by the fact that proteins are frequently phosphorylated on more than one type of amino acid and therefore also by different kinases.

These images were compared to the silver stained gel images and proteins of interest were pinpointed and identified (listed in Table 3).

## ***Limitations***

As mentioned in the introduction to this chapter, there are some drawbacks to this method. Antibodies can occasionally give non-specific signals, making accurate interpretation of the results difficult. There is a possibility that these antibodies recognise the same phosphorylation site on one protein, suggesting that the same protein has two distinct phosphorylation sites. Although more interference was expected between the **RXS** and **RXXS/T** antibodies, the blots of **TP** and **RXS** have more spots in common.

Antibodies can be highly specific and superimposition of the colloidal gold stained membranes with the blots and the analytical gels accurate (Liberatori et al. 2004). Even when this is achieved, it must be considered that the sensitivity of silver staining is lower than the sensitivity obtained by immuno staining. As a result low abundance proteins recognised by the antibodies can not be identified by MALDI-TOF MS because they are hardly visible on the silver stained analytical gels or overlap with more abundant proteins. An example of such a problem has been encountered with the spots 27, 28 and 18, 19 and 20. Although the signal from the immuno-blot is very clear and perfectly matched with the spots on the preparative gel, a mixture of two proteins was found in the same spots, making it impossible to determine which protein is binding to the antibody.

A different problem was encountered for spots 35, 36 and 38: although the immuno signal was strong and seemed very specific, after superimposition to pinpoint the corresponding protein on the analytical gel, the amount of these proteins was found to be below the sensitivity of the silver staining.

Nevertheless, in most other cases a unique match and identification was possible, and the proteins obtained were analysed. The sequence of each protein was also screened to find the phospho-motifs recognised by the antibodies.

### ***Substrates identified are involved in chaperone-redox network***

One of the limitations of this experiment is that only a small subset, mainly constituted by abundant proteins, could be identified with confidence, leaving out the low abundant signalling proteins. However, as this problem is widely encountered, a wealth of proteomics data on such proteins has been generated. This is a good starting point to further understand the role of high abundance proteins, which also appear the most versatile. The subset of potential kinases substrates is too limited to unveil possible connections within signalling cascades. However, some correlations with chaperone-regulated redox networks and its possible cross-talk to signalling are discussed in this chapter.

### ***Excluding Vimentin***

Vimentin was a common substrate for the three kinases. This protein followed the same phosphorylation behaviour as substrate of the three kinases, although it does not have RXS and TP consensus motifs in its sequence. Furthermore, Vimentin is the only substrate that PKA has in common with the other kinases. Combined with the fact that Vimentin is always found in mixture with other proteins this strongly suggests that is a false positive. In addition, Vimentin would be an ideal target for non-specific binding as it is abundant in the cell and is highly phosphorylated.



### *Chaperones and co-chaperones as a kinase substrate*

The general behaviour observed for Hsp8 is dephosphorylation on TP with exceptions for the isoform contained in spot 40. The isoforms which become immediately dephosphorylated before two minutes of EGF stimulation are the most acidic isoforms and therefore likely the most phosphorylated. The proteins at higher pI follow the same process at a slower rate. As for the isoform in spot 40, the peptide distribution strongly suggests that this could be a truncated form, possibly a degradation product. It is not possible to judge if it was a non-specific degradation that occurred during sample preparation, or a more specific product of a regulated proteolytic process that occurred in the cells. Interestingly, if indeed this was a proteosomic event, the increased phosphorylation might have a role in controlling the process.

As for phosphorylation on RXS, the Hsp8 isoforms were comparably not influenced by the stimulation, and no changes were observed. The isoform in spot 40 was not recognised by this antibody, suggesting that RXS might be located in the C-terminal of the protein.

It has been shown that Hsc70 forms complexes with Hsc90 and Casein Kinase II in resting platelets (Gear et al. 1997). Upon treatment with thrombin, Hsc70 dephosphorylation and dissociation from the complex occur, in agreement with these results. Investigations into these phosphorylation events included a 2D display of the isoforms, showing that only a subset of Stress induced phosphoprotein 1 isoforms become phosphorylated after 10 minutes of heat stress (Lassle et al. 1997). This is also in agreement with the data presented here: only two isoforms were immunostained with antibodies against phosphorylated TP and RXS, and only the isoform in spot 14 (with lower level of phosphorylation than the isoform in spot 13) showed a phosphorylation increase on TP and a decrease on KXS upon EGF stimulation. Thr198 is indeed in a TP consensus motif, confirming these findings. In addition to this, the consensus motif of MAPKAP kinase 2 is RXXS, which is the same as that recognised by the

antibody against PKA substrates. The data presented confirmed that Stress induced phosphoprotein 1 is not phosphorylated on sites which lie in this consensus. PDI becomes dephosphorylated upon stimulation, and a potential phosphorylation site is Ser150. An increasing phosphorylation of BiP upon EGF stimulation on possible RXXS/T consensus was also observed.

### *Redox-signalling crosstalk networks*

Oxidative conditions, as well as growth factors stimulation, are known to stimulate tyrosine phosphorylation of EGFR, leading to activation of ERK1/2 in a number of cell types (Droge 2002). Recent works have shed some light on how the activity of the chaperones is coordinated and revealed links between the stress induced response and the thermal response (Jacob et al. 2004). The following paragraphs describe the relationships between the proteins identified in this study and redox networks.

Interestingly, at least four of the chaperones identified in this work, namely Hsp8, BiP, PDI and GST-P, are protagonists in redox networks regulated by chaperones (Sitia et al. 2004). Hsp8 and Stress induced phosphoprotein 1 are protagonists in protein folding in the cytosol and nucleus, whilst BiP, PDI and GST play a crucial role in the second protein folding compartment in the cell, the endoplasmic reticulum (ER). Two distinct redox networks in the cytosol and ER guarantee that protein folding is strictly regulated, both under physiological and stress conditions. Mechanistically, the cysteines are the main targets as their thiol groups can be modified according to the redox conditions. When a disulfide bond (SS) is created by two reduced SH groups, a strong structural constriction is imposed on the protein, which can regulate its function. Cysteines behave like real reversible molecular switches and their controlled oxidation modulates signalling pathways, in a similar manner to phosphorylation.

Hsp8 and BiP are part of complexes (in the cytosol and in the ER respectively) and are able to fold polypeptides chains correctly and are sensitive to H<sub>2</sub>O<sub>2</sub> treatment (Baty et al. 2005).

Remarkably, Enolase-1 and phosphoglycerate mutase B (PGM), not previously connected with redox networks, have also recently been found to be sensitive to H<sub>2</sub>O<sub>2</sub> treatment (Baty et al. 2005). Enolase-1 was also found to be glutathionylated under resting conditions, suggesting that this protein can also be important as a glutathione store (Fratelli et al. 2002). Another observation is that upon oxidative stress only one isoform of Enolase-1 was previously found (Baty et al. 2005). Three isoforms have been identified here, only two of which are phosphorylated following similar kinetics. PGM has been identified as a new substrate for histidine kinase activity *in vitro* (Engel et al. 2004), whilst in this work it is strongly dephosphorylated on serine/threonine upon stimulation with EGF.

The mitochondrial transmembrane GTPase is involved in mitochondrial fusion, and stimulates the mitochondrial oxidation of substrates, cell respiration and mitochondrial membrane potential suggesting that this protein may play an important role in oxidation processes (Zorzano et al. 2004). Furthermore, mitochondrial transmembrane GTPase is a powerful regulator of cell proliferation via inhibition of ERK1/2 signal pathway (Chen et al. 2004). This protein is increasingly phosphorylated on TP, whilst the phosphorylation on RXS is not altered upon EGF treatment. Unfortunately, no further information on the exact phosphorylation sites could be obtained.

PDI creates disulfide bonds in the polypeptide chain, thereby reducing its own cysteines; additionally, it functions as isomerase to rearrange the disulfide bonds created until the protein is properly folded. When required, it can function as an unfoldase. PDI has long been known to be a phosphoprotein capable of autophosphorylation, and the phosphorylated region was shown to be within 146-306 (although the authors suggest that the phosphorylation does not

influence the activity) (Quemeneur et al. 1994). In the presented study PDI was dephosphorylated upon stimulation; a predicted phosphorylation site for this is Ser150.

As previously stated, BiP belongs to the family of Hsc70/Hsp70, but is found in the ER. The form found in this analysis showed a sensibly lower MW, due to degradation or different splicing. Early studies have shown this protein to be phosphorylated (but not autophosphorylated) on serine/threonine upon stress, with the phosphorylation causing dissociation from its substrates (Hendershot et al. 1988). Recent work has shown that during ER stress, malformed proteins accumulate in the ER, and BiP is dissociated from different cytosolic kinase domains leading to their activation (Shen et al. 2002). The ATPase activity of BiP is also thought to be coupled to detachment of substrates from PDI. The results shown demonstrate that there is an increasing phosphorylation upon EGF stimulation on a possible RXXS/T consensus site.

GST-P has different functions: it is involved in oxidant metabolism and regulation by catalyzing the glutathionylation of peroxiredoxin (Manevich et al. 2004), and has also been shown to bind to JNK and dissociate upon oxidative stress allowing JNK phosphorylation and activation (Adler et al. 1999). Not much has been published on GST-P phosphorylation, as the main regulatory mechanism for this protein is oxidation; the present results also suggest that no major changes in phosphorylation occur under the conditions for stimulation used here.

A more detailed and quantitative analysis of some of the potential substrates described in this chapter is offered in the following section.

## **CHAPTER 5**

### **PHOSPHO-SILAC STRATEGY**

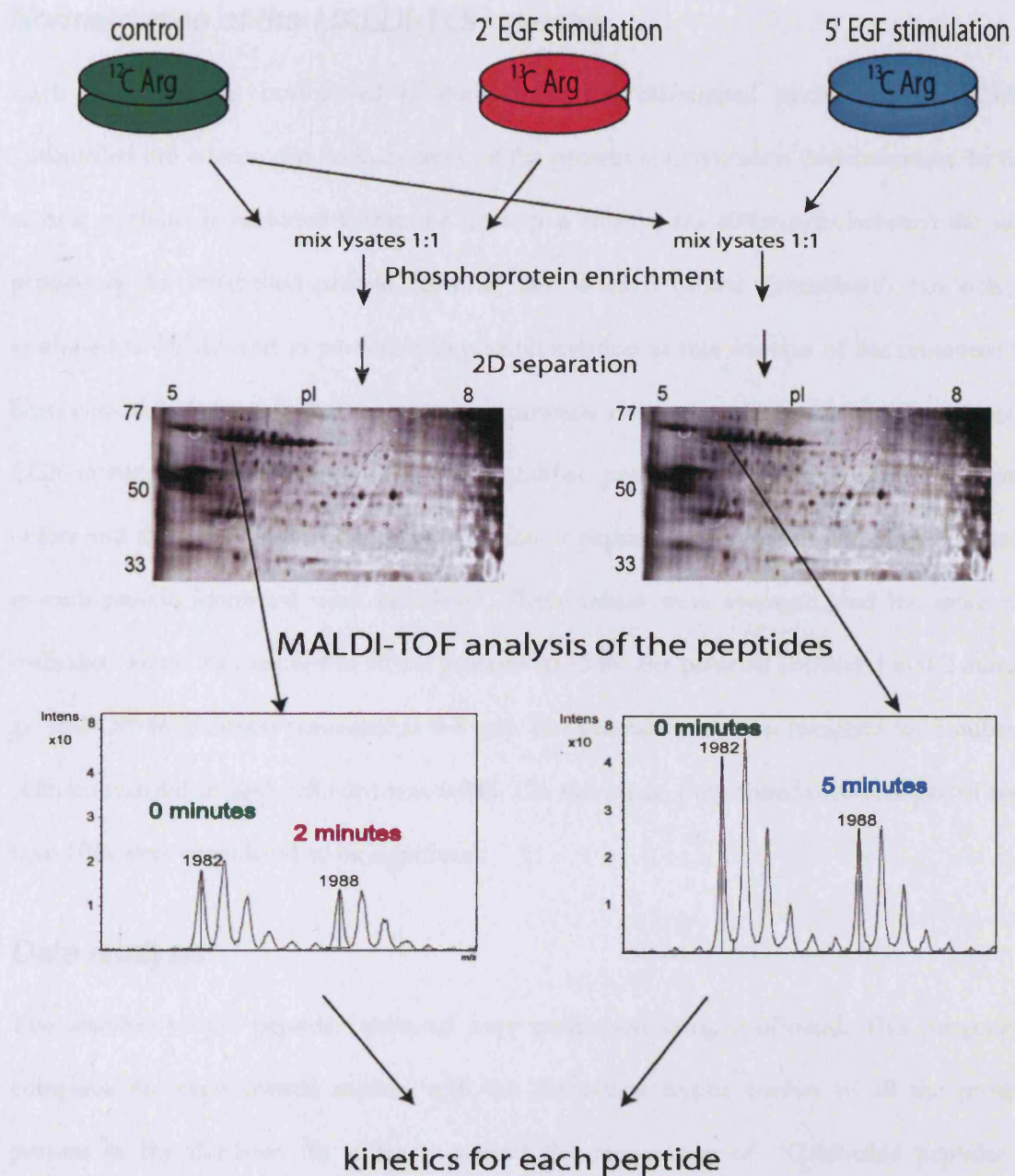
#### **INTRODUCTION**

A SILAC-based proteomics method for protein quantification was applied in the present EGF signalling study. In this approach, one population of cells was grown in medium containing the normal form of an essential amino acid and another was grown in medium supplemented with a stable isotope-labelled analogue (see chapter 1 for more details). The two resulting protein samples were then pooled and analysed. Fully labelled arginine containing six  $^{13}\text{C}$  atoms was chosen because it allows 6Da difference in weight between the labelled and the unlabelled peptide containing 1 arginine residue. This guarantees good separation of the corresponding peaks, and minimises the overlapping of the two isotopic patterns. The protease used for digesting the proteins was trypsin, which hydrolyses after arginine and lysine (equally abundant in proteins); this ensures that at least half of the peptides obtained contain arginine and can be quantified.

In this chapter a new strategy based on the combination of a more innovative application of SILAC and the previously described methods for selective enrichment of phosphoproteins (chapter 3) (Guerrera et al. 2005b) is described. This strategy allows quantitative, time-dependent scanning for phosphorylation for a majority of peptides containing arginine in many proteins identified on 2D gels. This use of SILAC also enables detection of changes in post-translational modifications other than phosphorylation of a certain protein upon stimulation of kidney fibroblasts with EGF. This suggests that this method is useful in characterising the correlation between different post-translational modifications in the same proteins or in different isoforms of the same protein.

The principal aim of the study described in this chapter is to investigate more thoroughly some of the proteins identified as potential targets of selected protein kinase families, as described in chapter 4. The Phospho-SILAC labelling was applied to phosphoproteomes of NRK49F cells stimulated for 2 and 5 minutes with EGF. The proteins of interest were isolated and analysed in detail.

This method proved ideal in integrating the analysis of previously functionally characterised proteins, and is equally applicable in obtaining interesting leads when analysing the whole phosphoproteome, as demonstrated in the following chapter.



**Figure 31. Scheme of the Phospho-SILAC strategy**

Cells were grown in labelled arginine containing media (pink and blue) and in ordinary media (green). After incorporation of labelled arginine the cells were treated with 200ng/ml EGF for 2 (pink) and 5 (blue) minutes. The cell lysates obtained from the control and each of the stimulated cell lines were mixed 1:1. Phosphoprotein fractions were isolated from both mixtures and separated on 2D gels. Each spot on the gel contained quantitative information for the resolved protein. The same spot from the two gels enclosed the data of the specific protein before and after 2 and 5 minutes of stimulation with EGF.

### *Normalisation of the MALDI-TOF spectra*

Each spectrum was normalised to compensate for differential loading of the labelled /unlabelled cell lysates, due to inaccuracy of the protein concentration determination. In fact, as new synthesis is excluded within the time span chosen, the differences between the same peptide in the unlabelled protein (control) and labelled protein (stimulated) can only be attributed to PTMs, and in particular to phosphorylation as this fraction of the proteome has been enriched. Assuming that no protein expression changes occur in the first 5 minutes of EGF treatment, the same intensities of unmodified peptides belonging to the same protein before and after stimulation is expected. Ratios of peptides attributed to unmodified peptides in each protein identified were calculated. These values were averaged, and the same normalisation factor was applied to all the proteins (0.53 for the proteins contained in 0-2 minutes gel and 1.45 for proteins contained in 0-5 gel). The standard deviation (weighed for number of pairs considered in each protein) was 0.088. On this basis, only abundance changes of more than 10% were considered to be significant.

### *Data analysis*

The searches of the peptides obtained were performed using ProFound. This programme compares the experimental masses with the theoretical tryptic masses of all the proteins present in the database. In order to permit the recognition of  $^{13}\text{C}$  labelled peptides an additional partial modification of +6Da was allowed on arginine. Potential partial phosphorylations were allowed on serine, threonine and tyrosine and partial oxidation was allowed on methionine. Allowance of a large number of partial modification is often associated to non-specific attribution of the peaks; therefore simple rules have been established with the purpose of limiting false interpretation:



- Only peaks that could not be attributed to the sequence were considered for interpretation of peptides with potential modification. Often a peak could be attributed to two or three different peptides on the same protein, and this trend increased when a large number of potential modifications were allowed. As phosphorylated peptides are more difficult to visualise on MALDI-TOF MS than unmodified peptides, it was assumed that if the same peak can be attributed to either a phosphopeptide or another peptide not carrying any modification, there is a higher possibility that the latter is the correct interpretation.
- When peaks were separated exactly by 6Da the interpretation of these was always given to the peptide containing modified (hR) and unmodified arginine (R). Even when it was not possible to recognise the peptides generating the doublet, alternative association of one of the two peaks to a peptide with PTMs was excluded.
- In most cases the attribution of  $^{13}\text{C}_6$  Arg containing peptides (especially for the ones carrying PTMs) was double checked to be not present on the MALDI-TOF spectrum of the same protein obtained from not labelled cells.

Although the application of these rules might have left out some potential modifications, it was necessary in order to exclude a large number of incorrect interpretations. A good example of the application of these criteria on a real search result is shown in Figure 32.

**g[38181549]gb[AAH61547.1] Heat shock protein 8 [Rattus norvegicus]**

Sample ID : 8  
 Measured peptides : 59  
 Matched peptides : 38  
 Min. sequence coverage: 44%

Measured Mass (M)	Avg/ Mono	Computed Mass	Error (ppm)	Residues	Missed	Cut	Peptide sequence
-------------------	-----------	---------------	-------------	----------	--------	-----	------------------

1196.649	M	1196.655	-5	459	469	0	FELTGIPPAPR
1202.709	M	1202.702	6	459	469	0	FELTGIPPAPR (1)+H6@R;
1198.671	M	1198.666	4	160	171	0	DAGTIAGLNVLRL
1204.739	M	1204.713	21	160	171	0	DAGTIAGLNVLRL (1)+H6@R;
1204.739	M	1204.519	182	518	526	1	<del>MVQEAEEKYK (1)+HPO3@STY;</del>
1216.730	M	1216.629	83	551	561	1	ATVEDEKLQKG
1252.663	M	1252.608	44	302	311	0	FEELNADLFR
1258.686	M	1258.655	24	302	311	0	FEELNADLFR (1)+H6@R;
1302.660	M	1302.591	53	540	550	0	NSLESYAFNMK
1318.629	M	1318.586	33	540	550	0	NSLESYAFNMK (1)+O@M;
1449.729	M	1449.797	-47	346	357	1	IQKLLQDFFNGK
1479.801	M	1479.746	37	300	311	1	ARFEELNADLFR
1491.833	M	1491.840	-5	300	311	1	ARFEELNADLFR (2)+H6@R;
1486.834	M	1486.693	94	37	49	0	TTPSYVAFTDTER
1492.803	M	1492.740	42	37	49	0	TTPSYVAFTDTER (1)+H6@R;
1491.833	M	1491.687	98	513	524	1	EDIERMVQEAEEK (1)+O@M;
1615.842	M	1615.780	39	113	126	0	SFYPEEVSSMVLTK
1631.823	M	1631.774	30	113	126	0	SFYPEEVSSMVLTK (1)+O@M;
1646.820	M	1646.626	118	37	49	0	TTPSYVAFTDTER (2)+HPO3@STY;
1648.840	M	1648.787	32	57	71	0	NQVAMNPTNTVFDAK
1664.815	M	1664.782	20	57	71	0	NQVAMNPTNTVFDAK (1)+O@M;
1658.928	M	1658.871	34	89	102	0	<del>HWPFMNVNDAGRPK (1)+H6@R;</del>
1658.928	M	1658.887	24	172	187	0	IINEPTAAAIAYGLDK
1690.782	M	1690.718	38	221	236	0	STAGDTHLGGEDFDNR
1696.803	M	1696.764	23	221	236	0	STAGDTHLGGEDFDNR (1)+H6@R;
1720.923	M	1720.698	131	329	342	0	SQIHDIIVLVGGSTR (3)+HPO3@STY;
1744.859	M	1744.801	33	584	597	1	NQTAEKEEFHQOK
1744.859	M	1744.748	63	57	71	0	NQVAMNPTNTVFDAK (1)+O@M; (1)+HPO3@STY;
1787.006	M	1786.982	13	172	188	1	IINEPTAAAIAYGLDKK
1787.006	M	1786.837	94	156	171	1	QATKDAGTIAGLNVLRL (2)+HPO3@STY;
1981.029	M	1980.990	20	138	155	0	TVTNNAVTVTPAYFNDSQR
1987.055	M	1987.037	9	138	155	0	TVTNNAVTVTPAYFNDSQR (1)+H6@R;
1981.029	M	1980.816	108	57	72	1	NQVAMNPTNTVFDAKRLxO@M; (2)+HPO3@STY;
1987.055	M	1986.863	97	57	72	1	NQVAMNPTNTVFDAKRLxO@M; (2)+HPO3@STY; (1)+H6@R;
1987.055	M	1986.871	93	113	128	1	SFYPEEVSSMVLTKMK2xO@M; (1)+HPO3@STY;
2259.128	M	2259.137	-4	362	384	0	SINPDEAVAYGAAVQAAILSGDK
2262.104	M	2262.109	-2	4	25	0	GPAVGIDLGTTYSCVGVFQHGK
2321.070	M	2321.058	5	37	56	1	TTPSYVAFTDTERLIGDAAK 2xHPO3@STY; (1)+H6@R;
2513.262	M	2513.300	-15	470	493	0	GVPQIEVTFDIDANGILNVSVDK
2773.232	M	2773.318	-31	424	447	0	QTQTFTTYSQNPQGVLIQVYGER
2779.245	M	2779.365	-43	424	447	0	QTQTFTTYSQNPQGVLIQVYGER (1)+H6@R;
2996.296	M	2996.449	-51	273	299	0	TLSSSTQASIEIDSLYEGIDFYTSITR
3002.290	M	3002.496	-68	273	299	0	TLSSSTQASIEIDSLYEGIDFYTSITR (1)+H6@R;

Excluded because assigned to part of a doublet AND because the alternative is an unmodified peptide

Excluded because assigned to an unmodified peptide

Excluded because assigned to unmodified doublet

**Unmatched Monoisotopic Masses:**

1211.726 1315.649 1457.875 1466.875 1537.927 1549.789 1554.775 1566.883 1568.849 1572.811 1601.790 1613.679 1753.719 1759.738 1761.748 1974.100 2043.998 2050.020 2175.060 2230.159 2325.043

**Figure 32. Example of selection of relevant attributions of a MALDI-TOF spectrum**

This figure represents the typical result page of the search engine most commonly used in the analysis, PROFOUND. Some of the attributions are excluded and a small explanation is given. The modifications are highlighted in red: H6@R stands for  $^{13}\text{C}_6$  Arg label (hR) and HPO3@STY stand for phosphorylation on serine, threonine or tyrosine.

## *Quantitation*

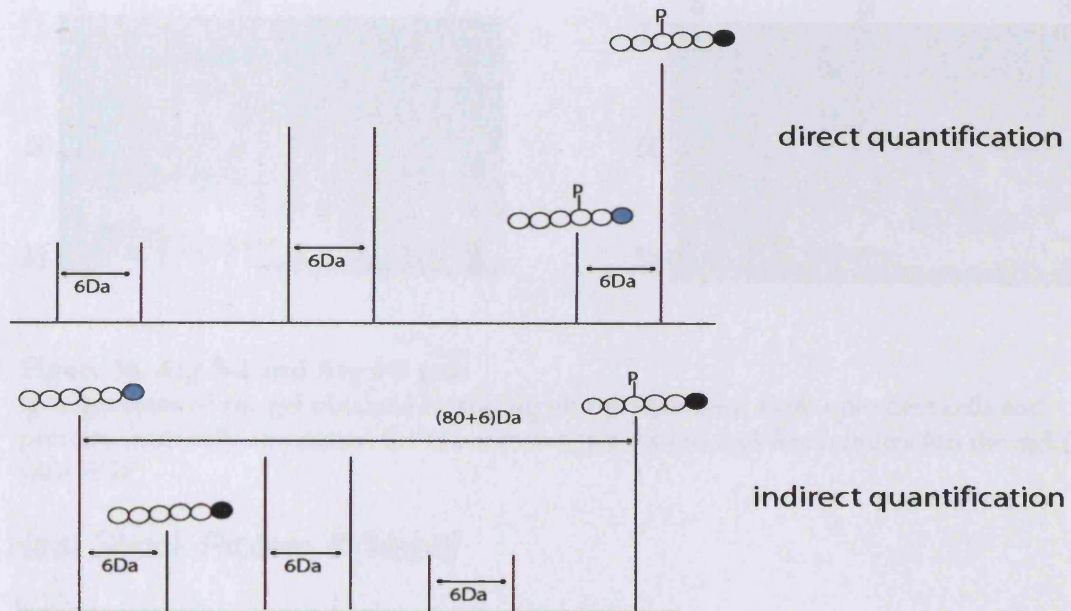
Strict rules were applied to quantification of the peptide doublets. First, special emphasis needs to be put on the fact that only intensities of peaks corresponding to structurally identical peptides can be compared. When one peptide appeared to be modified only in one of the analysed cellular states, no quantitative information could be obtained directly from this peptide, except on the on-off presence of such modification. When the corresponding unmodified peptide was detectable as a doublet some indirect conclusions could be drawn, because changes in the unmodified peptide must be complementary to changes in the corresponding modified peptide. A simple scheme is reported in Figure 33.

Some cases when the quantification within the same spectrum could not be obtained have been recognised:

- The peptide analysed did not contain any arginine;
- The peptide contained methionine both oxidised and unmodified and it was not possible to calculate accurately the amount of that peptide at different cellular states;
- On rare occasions, one of the peaks of the doublets of the peptide to quantify overlapped with the isotopic pattern of another peak.
- Finally, the comparison of quantification results obtained from different spectra of the same isoform of the same protein to obtain the kinetics of a certain phosphorylation at 0, 2 and 5 minutes were considered valid only when comparable coverage was obtained for the proteins analysed. For instance, this applies to results related to a spot from a gel containing a mixture of control and proteins from cells stimulated for 2 minutes with EGF (Arg 0-2 gel) and the same spot from a gel containing a mixture of control



and proteins from cells stimulated for 5 minutes with EGF (Arg 0-5 gel). The same rules applied when comparing the kinetics of different isoforms of the same protein.



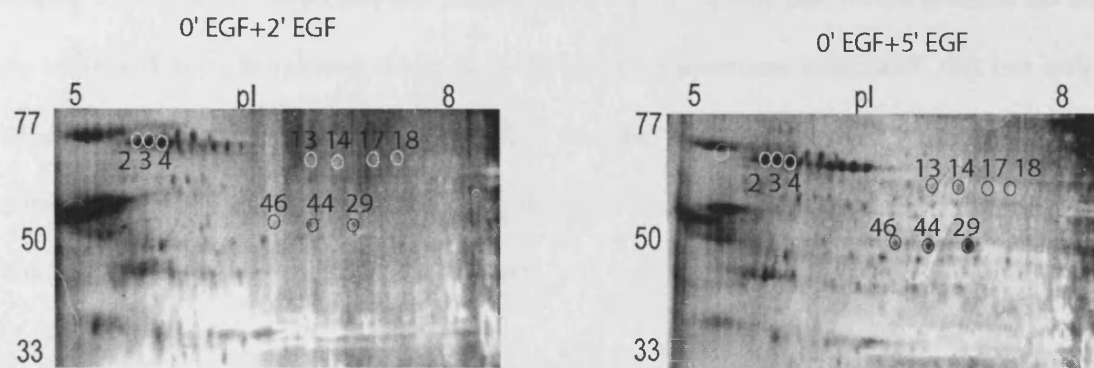
**Figure 33. Scheme of peptide quantification**

Black circles indicate heavy arginine, light blue circles indicate unlabelled arginine, empty circle indicate all other amino acids. P indicates phosphorylation. In the upper spectrum an example of phosphorylated peptide present before and after stimulation is reported. In this case quantification can be easily obtained calculating the ratio between the heights of the two peaks. In the lower spectrum a different case is reported, where the phosphorylation occurs only after stimulation. The ratio of the unmodified peptide before and after stimulation indicates the fraction of protein that undergoes phosphorylation upon stimulation.

## Protein kinase targets: Hsp8, Stress induced phosphoprotein 1 and Enolase-1

The sequence of different isoforms of the most abundant proteins found as kinase targets (as described in chapter 4) was investigated to potentially associate the phosphorylation found with the differential phospho-kinetics shown. Only spots that could be uniquely overlapped

with the gel reported in Figure 20 were included in the analysis. These spots, and the ones which were found to contain other isoforms of the same proteins, are shown in Figure 34.



**Figure 34. Arg 0-2 and Arg 0-5 gels**

Enlargements of the gel obtained by mixing phosphoproteins from quiescent cells and proteins from cells stimulated for two minutes (on the left) and five minutes (on the right) with EGF.

### *Heat Shock Protein 8 (Hsp8)*

#### *MALDI-TOF MS analysis*

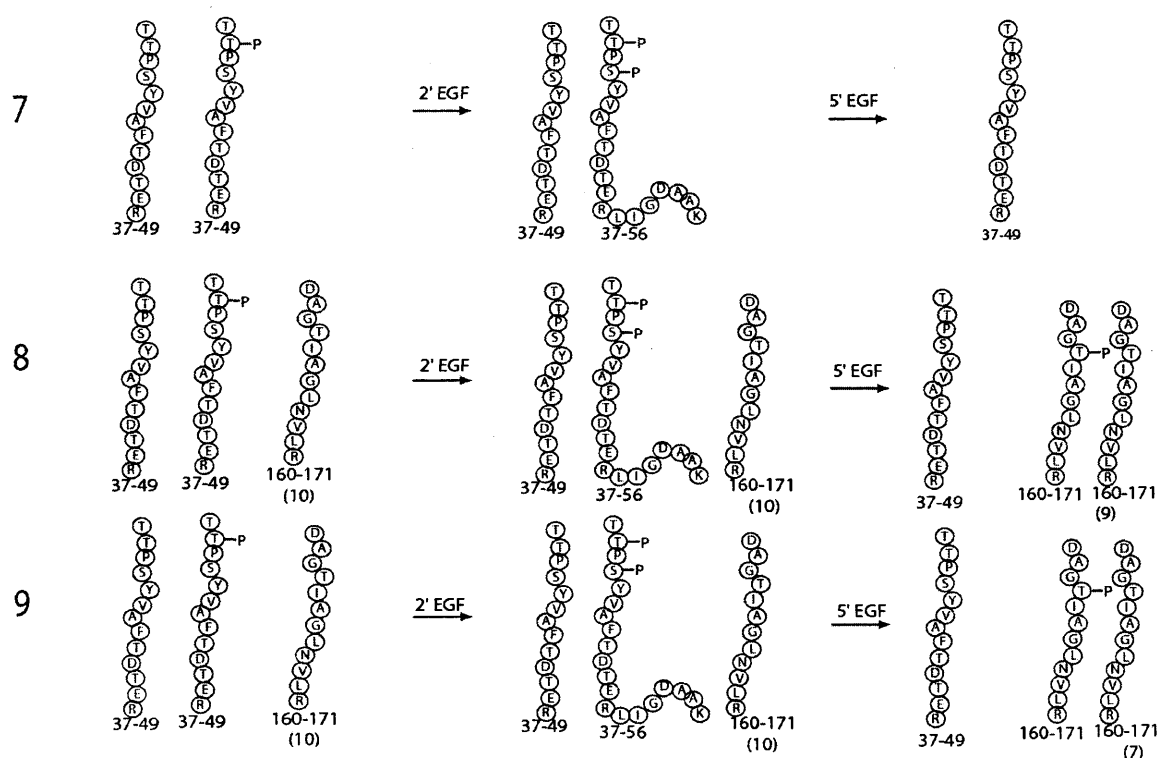
Spot 2, 3 and 4 corresponded to spots previously identified as Hsp8 (see chapter 4). Spots 1, 68 and 40 were not cut due to low abundance for spot 1 and 68 and poor co-localisation of spot 40. The coverage of the protein sequence using MALDI-TOF MS ranged from 40 to 65%, due to the abundance of the proteins contained in the spots. On one hand this constituted an advantage for this study, but on the other hand many potential phosphopeptides were excluded from the analysis as the above mentioned rules were followed, which implies that only the most probable sites are reported.

The results obtained are summarised in Figure 35; the isoform in spot 7 contained the peptide 37-49 both unmodified and monophosphorylated in resting cells. Upon stimulation there was an additional phosphorylation after stimulating for two minutes and complete dephosphorylation after 5 minutes. This is in agreement with the data shown in chapter 4: one

of the potential sites in this peptide is Thr38, which is in a MAPK substrate consensus sequence (TP). Unique attribution of the phosphorylation site was obtained by ESI MS (see below). The other isoforms showed the same behaviour for this part of the protein, but show an additional phosphorylation event after five minutes treatment with EGF; this can only be allocated to Thr163 (peptide 160-171hR-P, at 1285.7 m/z), which is not in any of the consensus sequences tested with antibody, therefore showed no signal on the western blot, but has been purified through the IMAC column.

Indirect quantification of the phosphorylation on Thr38 was hardly possible. The peak at 1493.7 m/z corresponding to 37-49hR was overlapping with the isotopic pattern of the peptide 300-311 2hR (1492.8 m/z).

Phosphorylation on Thr163 appeared to happen at different rates. According to the data obtained for the corresponding unmodified peptide 160-171, only 10% of the protein isoform 8 present in the cells became phosphorylated after 5 minutes, whilst up to 30% of the protein isoform 9 underwent phosphorylation after the same treatment.



**Figure 35. Phospho-kinetics of Hsp8**

Distribution phosphorylation upon stimulation with EGF in different Hsp8 isoforms. Each “worm” represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as –P. The name of the specific isoform analysed is indicated on the left.

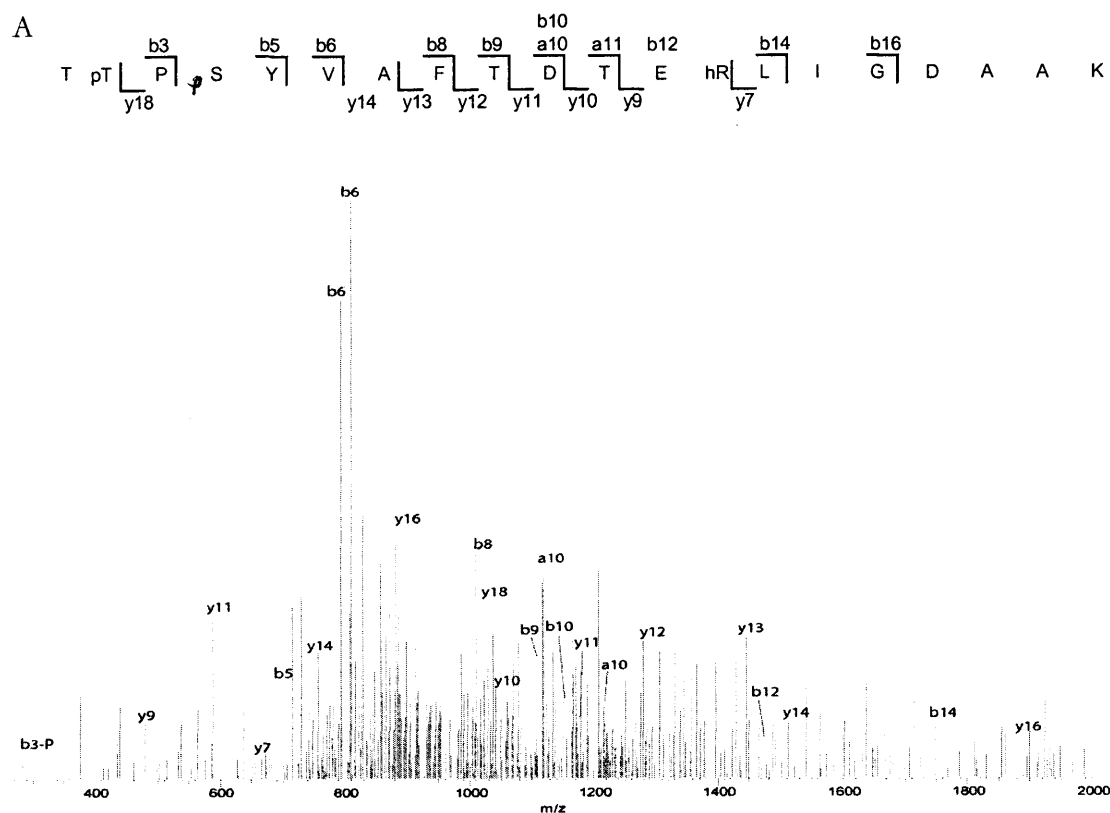
### *ESI MS analysis*

In order to uniquely allocate the phosphorylation sites ESI MS-MS analysis was performed on peptides 37-49 and 37-56. Peptide 160-171 only has only one possible phosphorylation site on Thr163, therefore it was not analysed.

Table 4 reports the ions that have been selected for tandem MS.

peptide	$^{12}\text{C}_6$ Arg	$^{13}\text{C}_6$ Arg	modifications	parent ion	charge	isolated ion
A 37-56		X	Diphosphorylation	2322.1	3+	774.7
B 37-49	X		Phosphorylation	1567.7	3+	523.2

**Table 4. Peptides selected for the ESI MS/MS analysis of Hsp8**

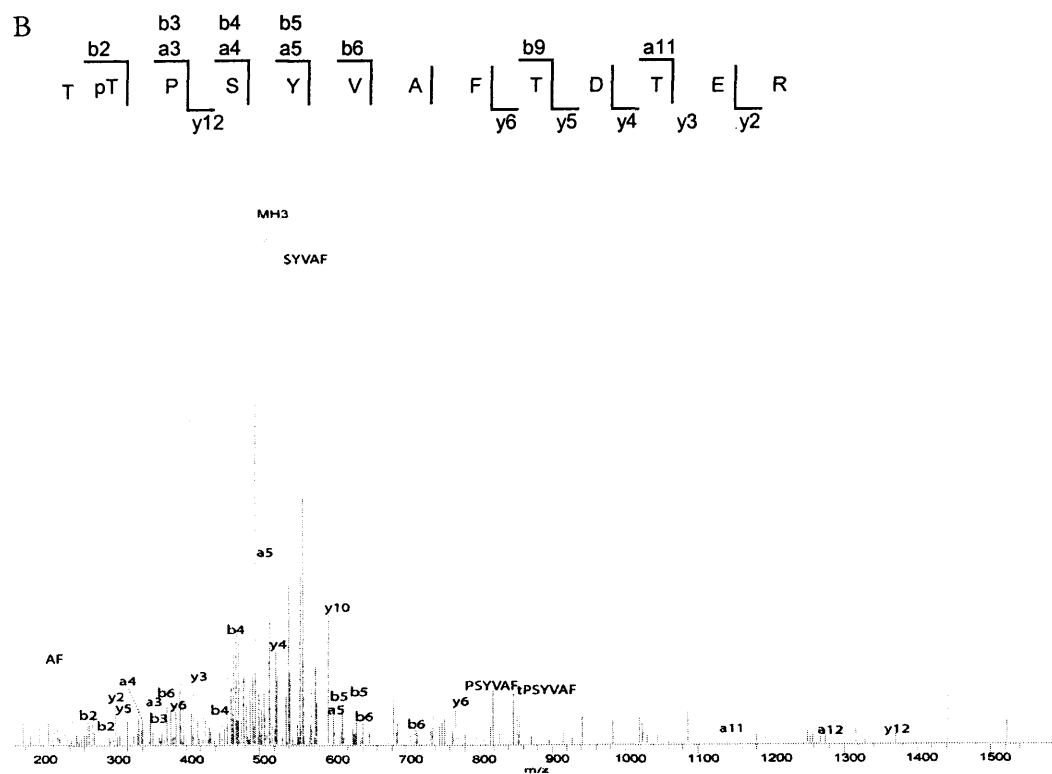


**Figure 36. ESI MS analysis of peptide A (as described in Table 4)**

The ions belonging to the peptide are annotated in the spectrum and they are listed in the table, with the theoretical values and the errors associated, expressed as difference in Da. On the top, a schematic representation of the peptide with the corresponding ions identified is reported. For convenience the charge and the  $\text{NH}_3/\text{H}_2\text{O}$  losses are not indicated on the spectrum and on the peptide scheme, but they can be found in the table. hR is  $^{13}\text{C}$  labelled Arg.

fragment	exper.	theor.	$\Delta$ Da
<b>a<sub>10</sub></b>	1215.10	1216.13	1.0
<b>a<sub>10</sub>-H<sub>3</sub>PO<sub>4</sub></b>	1118.90	1118.13	-0.8
<b>b<sub>3</sub>-H<sub>3</sub>PO<sub>4</sub></b>	282.80	282.32	-0.5
<b>b<sub>5</sub></b>	710.20	710.55	0.3
<b>b<sub>6</sub>-H<sub>2</sub>O</b>	792.30	791.67	-0.6
<b>b<sub>6</sub></b>	809.40	809.69	0.3
<b>b<sub>8</sub>-H<sub>2</sub>O</b>	1010.00	1009.93	-0.1
<b>b<sub>9</sub>-H<sub>2</sub>O</b>	1110.90	1111.03	0.1
<b>b<sub>10</sub>-H<sub>3</sub>PO<sub>4</sub></b>	1147.00	1146.14	-0.9
<b>b<sub>10</sub></b>	1245.20	1244.14	-1.1
<b>b<sub>12</sub></b>	1474.50	1474.36	-0.1
<b>b<sub>14</sub></b>	1747.80	1748.77	1.0
<b>b<sub>16</sub>-H<sub>2</sub>O</b>	1899.70	1900.97	1.3
<b>y<sub>7</sub>-NH<sub>3</sub></b>	671.10	670.79	-0.3
<b>y<sub>9</sub>-H<sub>2</sub>O<sup>+2</sup></b>	479.40	480.59	1.2
<b>y<sub>10</sub>-NH<sub>3</sub><sup>+2</sup></b>	880.80	880.77	0.0
<b>y<sub>10</sub>-H<sub>2</sub>O</b>	1061.20	1061.27	0.1
<b>y<sub>10</sub></b>	1078.20	1079.29	1.1
<b>y<sub>11</sub>-NH<sub>3</sub><sup>+2</sup></b>	589.40	589.18	-0.2
<b>y<sub>11</sub>-H<sub>2</sub>O</b>	1176.00	1176.36	0.4
<b>y<sub>12</sub>-H<sub>2</sub>O</b>	1277.90	1277.47	-0.4
<b>y<sub>12</sub></b>	1295.10	1295.49	0.4
<b>y<sub>13</sub></b>	1443.80	1442.66	-1.1
<b>y<sub>14</sub><sup>+2</sup></b>	756.30	757.38	1.1
<b>y<sub>14</sub></b>	1512.20	1513.74	1.5
<b>y<sub>18</sub><sup>+2</sup></b>	1020.50	1020.62	0.1
<b>y<sub>18</sub>-H<sub>2</sub>O<sup>+2</sup></b>	1011.20	1011.61	0.4





**Figure 37. ESI MS analysis of peptide B (as described in Table 4)**

The ions belonging to the peptide are annotated in the spectrum and they are listed in the table, with the theoretical values and the errors associated, expressed as difference in Da. On the top, a schematic representation of the peptide with the corresponding ions identified is reported. For convenience the charge and the  $\text{NH}_3/\text{H}_2\text{O}$  losses are not indicated on the spectrum and on the peptide scheme, but they can be found in the table.

fragment	exper.	theor.	$\Delta$ Da
<b>a<sub>3</sub></b>	352.6	352.31	-0.3
<b>a<sub>4</sub>-H<sub>3</sub>PO<sub>4</sub></b>	340.3	341.39	1.1
<b>a<sub>5</sub>-H<sub>3</sub>PO<sub>4</sub></b>	505.3	504.57	-0.7
<b>a<sub>5</sub></b>	602	602.56	0.6
<b>a<sub>11</sub>-H<sub>3</sub>PO<sub>4</sub></b>	1140.4	1139.26	-1.1
<b>b<sub>2</sub>-H<sub>2</sub>O</b>	265.2	265.18	0.0
<b>b<sub>2</sub></b>	282.5	283.2	0.7
<b>b<sub>3</sub>-H<sub>2</sub>O</b>	362	362.3	0.3
<b>b<sub>4</sub>-H<sub>2</sub>O</b>	448.9	449.38	0.5
<b>b<sub>4</sub></b>	468.4	467.4	-1.0
<b>b<sub>5</sub>-H<sub>2</sub>O</b>	613.7	612.56	-1.1
<b>b<sub>5</sub></b>	630.1	630.57	0.5
<b>b<sub>6</sub>-H<sub>3</sub>PO<sub>4</sub></b>	632.2	631.71	-0.5
<b>b<sub>6</sub>-H<sub>2</sub>O</b>	712.6	711.69	-0.9
<b>b<sub>9</sub>-H<sub>2</sub>O</b>	1030.4	1031.05	0.6
<b>y<sub>2</sub></b>	303.5	304.33	0.8
<b>y<sub>3</sub></b>	405.8	405.43	-0.4
<b>y<sub>3</sub><sup>+2</sup></b>	203.4	203.22	-0.2
<b>y<sub>4</sub></b>	521.6	520.52	-1.1
<b>y<sub>5</sub><sup>+2</sup></b>	311.3	311.32	0.0
<b>y<sub>6</sub><sup>+2</sup></b>	384.8	384.91	0.1
<b>y<sub>6</sub></b>	767.7	768.81	1.1
<b>y<sub>12</sub>-H<sub>3</sub>PO<sub>4</sub></b>	1370.5	1369.48	-1.0
<b>AF</b>	220.1	219.27	-0.8
<b>SYVAF-28</b>	540.4	540.64	0.2
<b>tPSYVAF-28</b>	818.9	818.85	0.0
<b>tPSYVAF</b>	846.1	846.86	0.8

## *Stress induced phosphoprotein 1*

No potential phosphorylation complied with all the rules described in the introduction of this section. However, it must be pointed out that only 2 isoforms of this protein were found as substrates of kinases, whilst an additional two isoforms at higher pI have been identified here, suggesting differential phosphorylation.

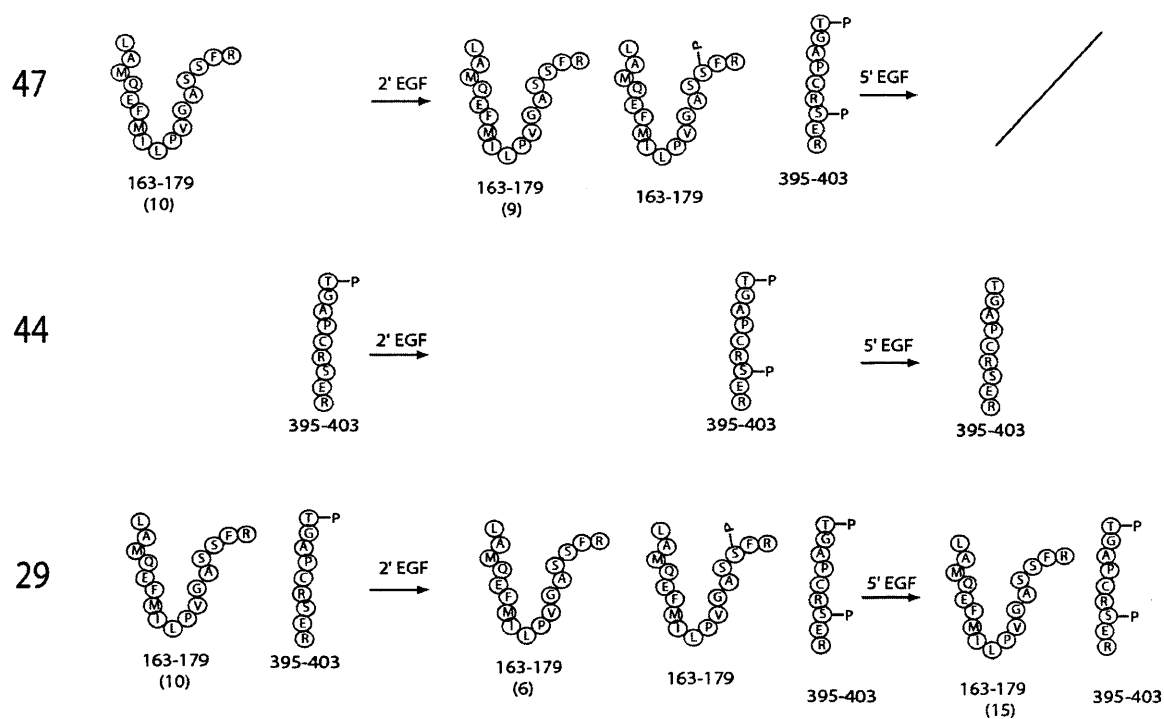
## *Enolase-1*

Spots described as 46 and 44 were both found in this analysis. In the previous experiments (chapter 4) spot 44 could not be identified as Enolase-1. Spot 29, which did not correspond to any spot from the kinase fishing experiment, was also found to contain Enolase-1. In total four spots containing Enolase-1 were found (spots 46, 29, 44 in Figure 34 and spot 49 in Figure 20): only spots 46 and 49 were recognised by the antibody against phosphorylated TP. Spot 46, recognised by the antibody and analysed in this SILAC experiment for time 0 and 2 minutes, showed phosphorylation on serine 177 (or Ser176) upon stimulation, but no phosphorylations on TP were visible. Spot 44 and spot 29 have been analysed in this SILAC experiment, but were not recognised by any of the antibodies used, and showed phosphorylation on serine/threonine. In detail, spot 47 could only be analysed for the first two minutes of stimulation, because the spectrum from the gel containing proteins from control and cells stimulated for 5 minutes could not successfully be taken. Enolase-1 in spot 44 was analysed throughout the stimulation, although a small reserve must be expressed for the peak corresponding to the peptide 395-403, with two arginines labelled and no other modifications. The isotopic profile of this peptide overlapped to the one of a strong peak deriving from auto digestion of trypsin, therefore it is likely to be covered by it. Finally Enolase-1 in spot 29 underwent phosphorylation on two different peptides, 395-403 which from monophosphorylated became biphosphorylated, and stayed so up to 5 minutes stimulation;

peptide 163-179 instead became transiently monophosphorylated at 2 minutes, only to lose it after 5 minutes of treatment with EGF. It would have been interesting to calculate the percentage of the protein that became phosphorylated, but this was not possible as no traces of phosphorylated peptide were present at 0 and 5 minutes. Indirect quantification was not possible because unmodified 163-179 was present as a mixture of peptides with different degrees of oxidised methionine. Spot 46 was not analysed.

The amino acids involved are not in a consensus sequence searched, consequently it is not surprising that they were not recognised by any of the antibodies. A scheme of the results obtained is shown in Figure 38.

The ratio between the peptide 163-179 before and after the stimulation of isoform 47 was near 1:1, suggesting that the portion of protein that undergoes this phosphorylation is very small. However, in isoform 29 as much as 40% became phosphorylated after 2 minutes. Inexplicably, the ratio between the unmodified peptide after 5 minutes of EGF treatment and the control is 3:2, suggesting that this peptide might be reversely modified in the control but such modifications could not be identified.



**Figure 38. Phospho-kinetics of Enolase-1**

Distribution phosphorylation upon stimulation with EGF in different Enolase-1 isoforms. Each “worm” represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as -P. The name of the specific isoform analysed is indicated on the left.

## **DISCUSSION**

The focus of the work presented in this chapter was on Hsp8 and Enolase-1 analysis. A variety of their phospho-isoforms was covered in an attempt to understand the role that they play in response to EGF stimulation. The observations on the structures reported in the following paragraphs have been made in collaboration with Dr. Nicholas Keep.

### **Hsp8 phosphorylation**

A large number of heat shock proteins, which constitutes a major class of proteins chaperones (described in chapter 1), were identified using the broader Phospho-SILAC approach. Due to the high sequence homology between Hsc70 and Hsp8 (more than 99%), considerations on function and structure reported for Hsp70/Hsc70 will be considered valid for Hsp8.

It has long been known that the family of Hsp70 undergoes autophosphorylation, and that the level of phosphorylation of some Hsp70 proteins *in vivo* is responsive to stress and other cellular conditions (Leustek et al. 1992). In some cases phosphorylation has been shown to increase the binding of Hsp70 substrates (Sherman et al. 1993). Only recently Hsp70 proteins have been shown to be phosphorylated by PKA and that this provides an on/off switch for the regulation of protein phosphatase 2B signalling by Hsp70 (Lakshmikuttyamma et al. 2004). No records of exact phosphorylation sites are available.

Results presented in chapter 4 showed that Hsp8 is a potential target for MAPK and PKC kinase families. In this chapter the protein Hsp8 isoforms and their phosphorylation kinetics in different phosphorylation sites were analysed in more detail. All isoforms identified in this part of the analysis (the three most abundant) are phosphorylated on Thr38 (TP motif) in quiescent cells. In addition, a second site of phosphorylation was found in the same domain. Ser40 only became phosphorylated within the first two minutes of stimulation, presumably to support the

role of the phosphorylation of Thr38. Both phosphorylations were removed after 5 minutes of EGF treatment. Unmodified proteins were present throughout the stimulation, suggesting that these changes only involved a fraction of the Hsp8 isoforms present in the cell. Unfortunately, it was not possible to quantify the fraction that became phosphorylated for each protein isoform, for two main reasons. First, identical phosphorylated peptides were not present in all Hsp8 spectra. Second, indirect quantification was not possible because the unmodified peptide peaks were overlapping with isotopic pattern of other peaks. Despite the limitations of the method encountered in this specific case, a high level of information was achieved: multiple Hsp8 isoforms were distinguished and different phosphorylation sites and relative kinetics were separately studied for each isoform.

An additional phosphorylation has been found on Thr163, which is not in any of the substrate consensus screened in the previous experiment (see chapter 4). This residue became phosphorylated only after 5 minutes of stimulation with EGF. Interestingly, this activation seems isoform-dependent. Hsp8 isoform 2 did not become phosphorylated, one in ten molecules of Hsp8 isoform 3 became phosphorylated on Thr163 after 5 minutes, and up to three in ten molecules of Hsp8 isoform 4 did the same. Different isoforms seemed to become phosphorylated in a wave, maybe to more finely modulate cellular processes.

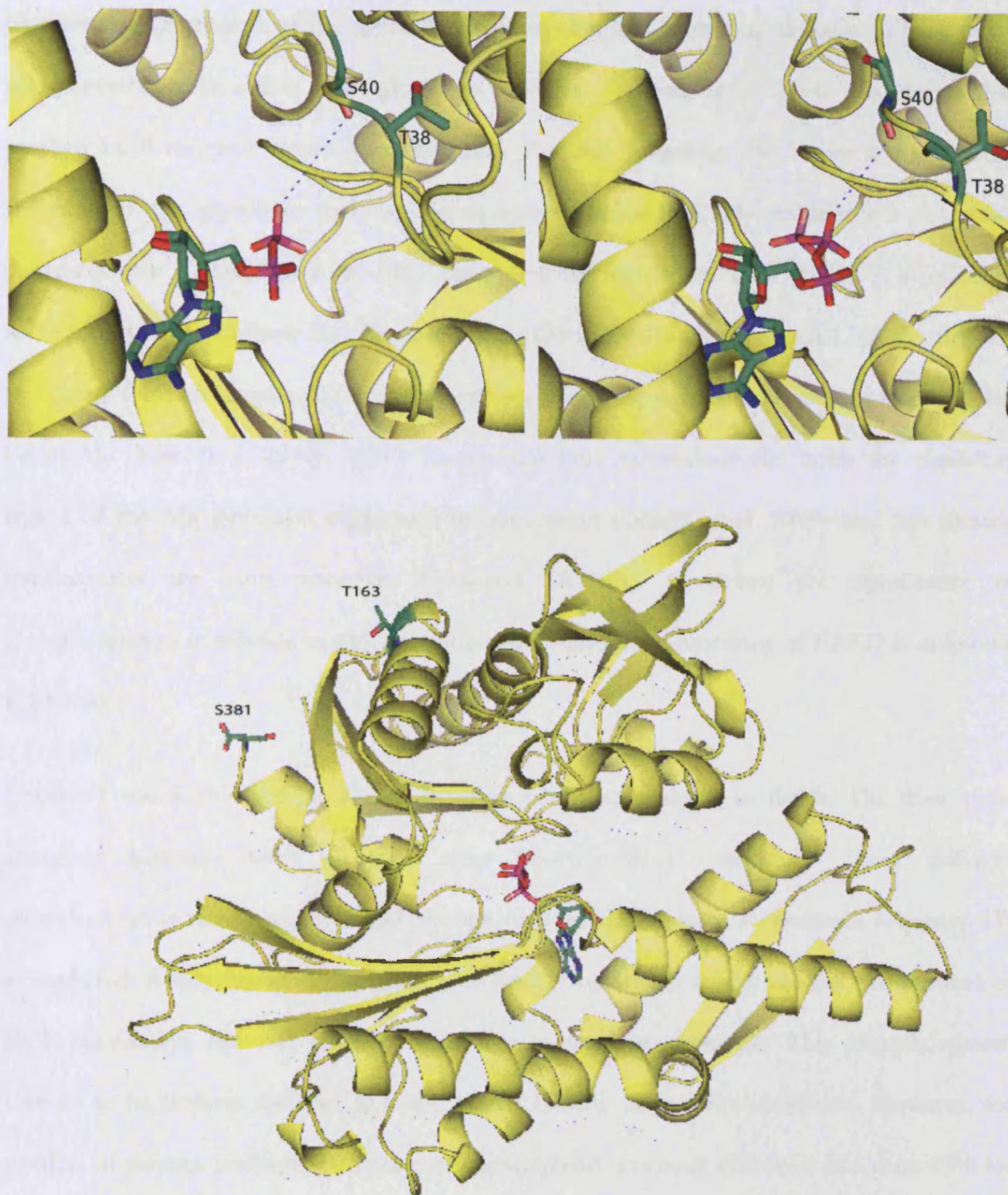
In order to understand the functional meaning of these multiple phosphorylations the presented study should be completed with molecular and structural biology experiments. However, very good leads could be obtained by observing the Hsc70 structure. The commentary on this structure has been made in collaboration with Dr. Marialuisa Pellegrini-Calace.

All the modifications found in this study fall in the N-terminal of the ATPase domain (refer to the introduction for more details on Hsc70/Hsp70 structure). Phosphorylations on Thr38 and

Ser40 might influence the nucleotide exchange rate by stabilising the binding with ADP. As shown in Figure 39, the ADP is surrounded by two loops which stabilise the binding between ADP and the Hsp8 via H bonds between the NH of the backbone and the phosphates. Phosphorylation on both Ser40 and Thr38 might push the loop towards the ADP molecule, stabilising the binding and therefore favouring Hsp8 chaperoning activity. However, the accessibility of these two peptides is very limited, and phosphorylation on these sites must imply a strong structural change (Figure 39). After five minutes of EGF treatment, these residues are dephosphorylated, possibly to allow the exchange of ADP for ATP, as the dephosphorylated form of the protein seems to favour the ATP binding. Interestingly, only a portion of the protein is phosphorylated on Thr38 in resting cells, and the same portion probably becomes diphosphorylated within the first two minutes of stimulation. This is likely due to the fact that Hsc70s are highly expressed and play many roles in the cell; partial phosphorylation might define the portion available for short term stresses.

Thr163 is on the surface of the ATPase domain, and its phosphorylation might influence binding with other proteins, or with the C-terminal part of the protein. It is not possible to analyse this further as a structure of the whole protein is not available.

In conclusion, the modifications on the N-terminal of Hsp8 might influence its activity by regulation of the ADP/ATP exchange. In particular, when Hsp8 is phosphorylated the ADP bound form is favoured and the chaperoning activity is increased. This would result in activation of the signalling proteins previously found to be associated with Hsp8. After a few minutes of EGF treatment the ATP form is stabilised by removal of the phosphorylation on Hsp8 and the substrate released in its active form.



**Figure 39. Hsc70 structure 1kax**

The amino acids of interest have been highlighted. Top panels are an enlargement of the site of the protein binding ADP (left) and ATP (right). Bottom panel shows the whole ATPase domain of Hsc70 and T163 as a potential phosphorylation site and Ser381 which is the C-terminal of the structure.



## Enolase-1 involvement in EGF signalling

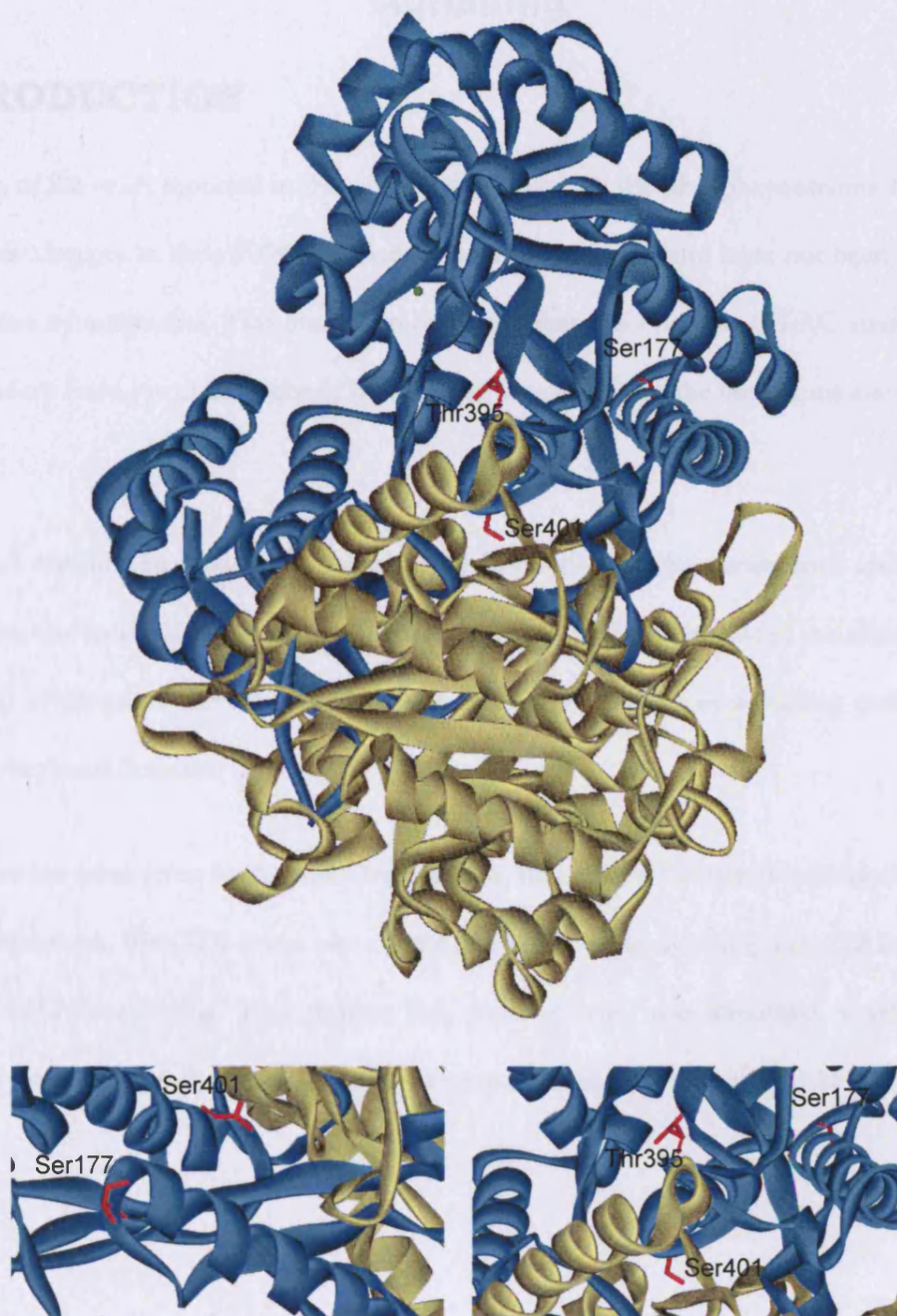
Enolase-1 is an enzyme of the glycolytic pathway, but has been long thought to have other roles, together with a few other glycolytic proteins. Twenty years ago it was shown that purified EGF receptor kinase phosphorylates Enolase-1 together with other key regulatory enzymes of the glycolytic pathway, phosphofructokinase and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), in an EGF-dependent manner (Reiss et al. 1986). A number of more recent works indicate that Enolase-1 (or MBP1, Myc-binding protein 1, an alternatively translated form of Enolase-1) is a transcriptional repressor of the oncogenic transcription factor *Myc* (Kim et al. 2004). MBP1 translocates into the nucleus and binds the regulatory region of the *Myc* gene and suppresses its expression (Ghosh et al. 1999) and *Myc* directly transactivates the gene encoding Enolase-1 (ENO1). However, the significance of phosphorylation in relation to the glycolytic or non-glycolytic functions of ENO1 is unknown at present.

Enolase-1 was highly abundant and therefore could be analysed in depth. The three most abundant isoforms were analysed after Phospho-SILAC isolation. Three different phosphorylation sites were identified, though none of them was in a consensus sequence TP, as expected. A transient phosphorylation on Ser177 or Ser176 appeared after two minutes of EGF stimulation and was removed after five minutes of treatment. This phosphorylation seemed to be isoform specific, as it was found on two of the three isoforms. However, the portion of protein isoforms becoming phosphorylated was very different: less than 10% for one Enolase-1 isoform and up to 40% for the other. Another phosphopeptide identified was 395-403; this peptide contained two phosphorylation sites, Thr395 and Ser401. Isoforms 47 and 29 were monophosphorylated in untreated cells, and they became diphosphorylated within two minutes of EGF treatment.

It is very difficult to make hypotheses on the role of the phosphorylations found in this study because they do not fall in any of the functional sites, such as metal binding site or protein interaction sites. However, the residues of interests were located on the structure from a sequence with 83% identities and 90% positives. This structure reports the whole protein as a homodimer (Chai et al. 2004) (Figure 40). Ser401 is pointing into the interface of the dimer, possibly regulating the affinity of the two subunits; in addition, overlapping the structures of the monomer and the homodimer shows that Ser401 is the amino acid that undergoes the strongest change. It is very difficult to determine the role of this residue during EGF stimulation; two isoforms out of three present a monophosphorylation in this region that could be attributed to either Ser401 or Thr395. Thr395 could easily accommodate a phosphate group, which could even stabilise the structure via a number of favourable neighbours. Nevertheless, this amino acid is quite inaccessible, which suggests that the phosphorylation might be structural. Unfortunately, it is not possible to confirm that this phosphorylation is constitutive because the structures have been determined only from proteins expressed in bacteria. It seems that Thr395 phosphorylation is less likely to be functional compared to the Ser401 phosphorylation, and that the latter is probably connected to a signalling event. It could indeed cause dimer dissociation, and inactivation as glycolytic enzyme. However, it was not possible to quantify the changes, leaving most questions open.

Ser177 (and not Ser176) is conserved in the protein structure; this residue is engaged in stabilising the end of the helix and seems to be very accessible, although its function in destabilising this part of the structure under EGF induced phosphorylation (in an isoform specific way) is hard to predict. It becomes phosphorylated in two isoforms only at two different rates, but is already dephosphorylated after five minutes, suggesting that this phosphorylation is a rapid effect of the early stage of the stimulation. In Figure 40 the structure the three residues of interest discussed above is displayed.

In conclusion, stimulation with EGF might cause dissociation of part of the Enolase-1 present in the cell, maybe in a negative regulation feedback, or as part of redox network regulation.



**Figure 40. Enolase-1 structure 1te6**

Top panel shows the whole structure with the amino acids of interest highlighted. Bottom panels are an enlargement of the site of the homodimer interface.

## **CHAPTER 6**

### **Phospho-SILAC strategy wider application in signalling**

#### **INTRODUCTION**

The aim of the work reported in this chapter was to screen the phosphoproteome for proteins that show changes in their PTMs in relation to the EGF signal and have not been highlighted by the use of antibodies. This study demonstrated that the Phospho-SILAC strategy can be independent from previous antibody based screenings and from the limitations associated with them.

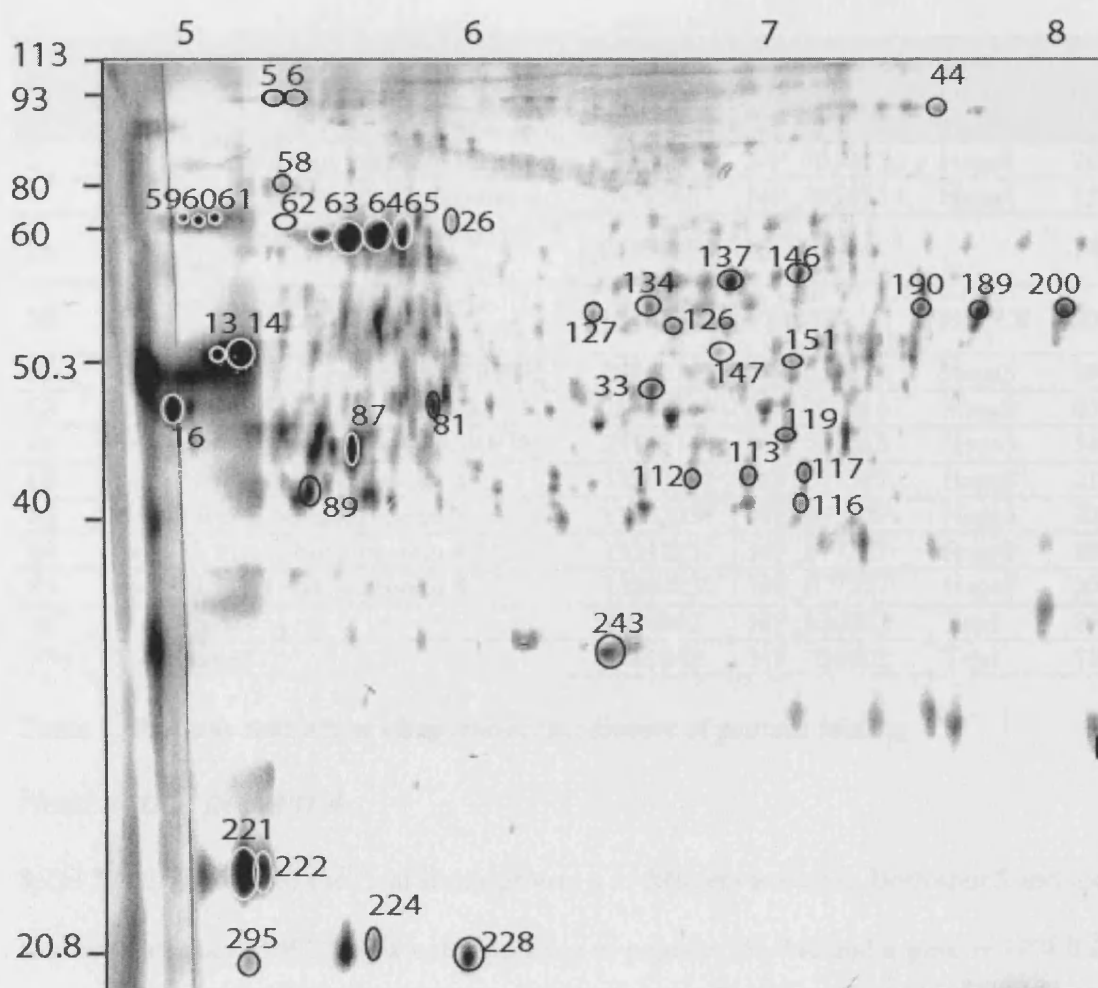
A 2D gel obtained by mixing the phosphoprotein fractions from unstimulated and stimulated cell lysate was randomly analysed and protein spot were excised, processed and identified. This provided evidence of the value of the Phospho-SILAC strategy in signalling studies to find leads to key modifications.

Attention has been given to the signalling proteins. Considered the nature and the dynamics of its modifications, RhoGDI-1 was one of the most interesting signalling proteins found in the present SILAC screening. Two distinct Rab proteins were also identified, together with a typically neuronal signalling protein, collapsin response mediator protein (CRMP-2).

## RESULTS

Proteins from the whole phosphoproteome were analysed by SILAC/phospho-enrichment combination. Three hundreds spots were cut from a gel which contained phosphoproteins from control cells and phosphoproteins from stimulated labelled cells. The analysis was performed only for the first two minutes of stimulation with EGF as it has been shown that the activation of MAPK cascade lies within this time (chapter 4).

Forty proteins were selected and divided in different categories according to their role in the cell. The analysis of the data obtained is presented according to the class of proteins identified. Due to constant changes of accession numbers (Gi No) in the used non-redundant database (NCBI) each protein in the tables is accompanied by Gi No, the closest reference sequence (RefSeq) and the name of the corresponding gene, when known.



**Figure 41. Arg 0-2 gel**

Enlargements of the gel obtained by mixing phosphoproteins from quiescent cells and proteins from cells stimulated for two minutes with EGF. Spots annotated were analysed.

### Other chaperones

A whole section of these results is dedicated to chaperones proteins in the attempt of understanding a possible common behaviour between them and Hsp8, previously analysed. 14 proteins belong to the class of chaperones proteins, and they are listed in Table 5. For spots 26, 58, 127 and 134 no differences between the protein coming from the stimulated cells and the control were observed.

spot	Proteins that act as chaperones/mediators of protein folding	Gi No	Refseq	Gene	Cov.
5	Heat shock protein 70kDa protein 4	24025637	NP_705893.1	Hspa4	28%
6	Heat shock protein 70kDa protein 4	24025637	NP_705893.1	Hspa4	15%
26	dnaK-type molecular chaperone grp75 precursor	2119726	I56581		14%
58	Heat shock protein HSP90-beta (HSP84)	1346320	P34058	HSPCB	20%
59	Heat shock 70kD protein 5 (GRP78)	25742763	NP_037215	Hspa5	28%
60	Heat shock 70kD protein 5(GRP78)	25742763	NP_037215	Hspa5	43%
61	Heat shock 70kD protein 5(GRP78)	25742763	NP_037215	Hspa5	36%
62	Hscp70. Heat shock protein 8	13242237	NP_077327	Hspa8	21%
63	Hscp70. Heat shock protein 8	13242237	NP_077327	Hspa8	22%
64	Hscp70. Heat shock protein 8	13242237	NP_077327	Hspa8	28%
65	Hscp70. Heat shock protein 8	13242237	NP_077327	Hspa8	20%
127	T-complex1	6981642	NP_036802	Tcp1	25%
134	T-complex1	6981642	NP_036802	Tcp1	15%

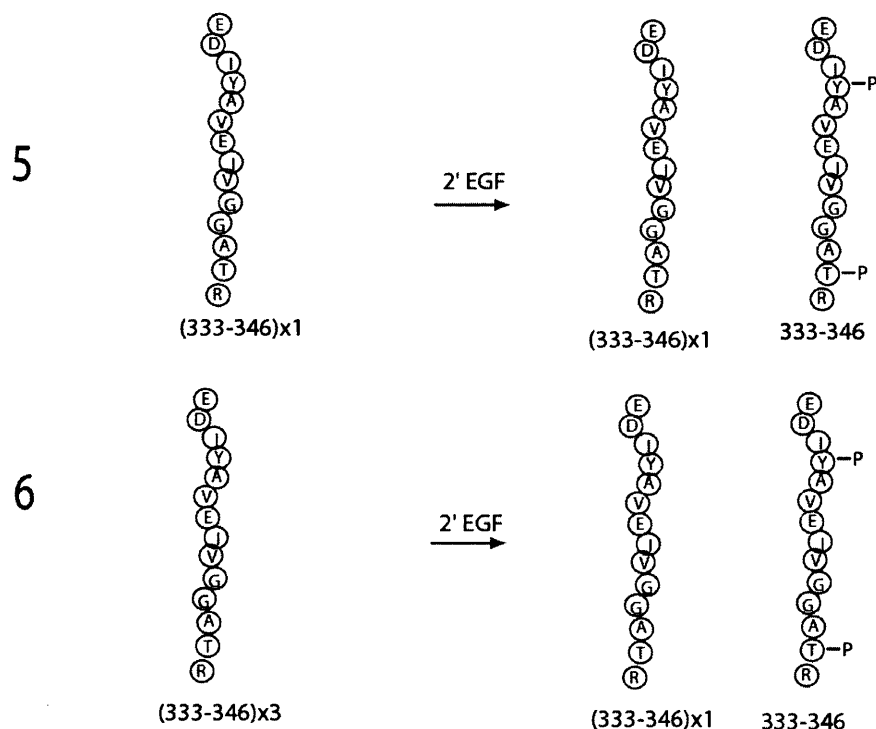
**Table 5. Proteins that act as chaperones/mediators of protein folding**

### *Heat shock protein 4*

Spots 5 and 6 contained the Heat shock protein 4 in different isoforms. Both spot 5 and spot 6 contained a peak at 1492.9 m/z corresponding to peptide 333-346 and a peak at 1498.9 m/z corresponding to the same peptide after stimulation. In addition, in both spectra there was a peak at 1658.8 m/z corresponding to the same peptide 333-346 with  $^{13}\text{C}_6$  Arg plus two phosphorylations. As the phosphorylated peptide was not present in this protein in unstimulated cells, the amounts of the unmodified peptide from same proteins in unstimulated and stimulated cells were compared instead. Interestingly, in spot 5 the ratio of the intensities of the peaks 1492.9 m/z and 1498.9 m/z was close to 1:1, suggesting that only few copies of the protein get phosphorylated, whilst in spot 6 it was as much as 30% (ratio of the intensities of the peaks 1492.9 and 1498.9 is 1.5). Furthermore, sample 5 contained a peak at 1757.8 and one at 1838.0 corresponding to peptides 170-185 and 170-185-P. Peptide 170-185 did not contain any arginine, therefore no conclusions could be made about the correlation between

stimulation and phosphorylation, but this might contribute to the slight shift towards the more acidic pH compared to spot 6.

A scheme of the different peptides observed in the two isoforms of Hsp4, as well as its quantitative model, is represented in Figure 42.



**Figure 42. Phospho-kinetics of Hsp4**

Distribution phosphorylation upon stimulation with EGF in different Hsp4 isoforms. Each “worm” represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as -P. The name of the specific isoform analysed is indicated on the left.

### *Heat shock protein 5*

Spots 59, 60, and 61 contained the Heat shock protein 5 in different isoforms. In this case as well, the respective MALDI-TOF spectra were almost identical. Peptides 377-386, 377-386hR (at 1210.6 m/z and 1216.6 m/z respectively) were present in the three spots. However, peptide 377-386hR-P (1296.7 m/z) was only present in spots 59 and 60; this suggests that

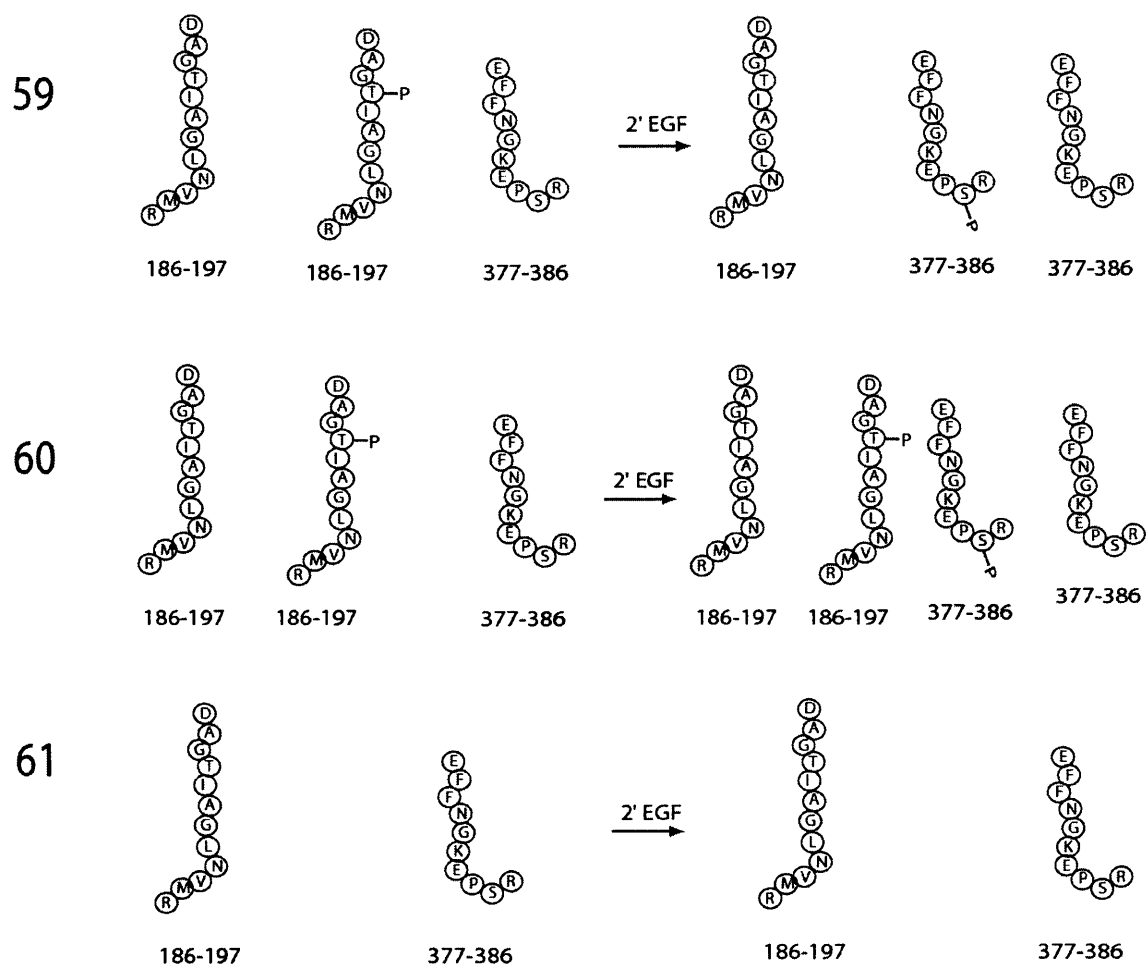


phosphorylation on Ser385 occurs selectively on some isoforms as a response to EGF stimulus.

The pair of peaks attributed to peptides 186-197 and 186-197hR (1217.6 m/z - 1223.6 m/z) were present in all three spots, whilst only the isoform in spot 59 contained 186-197 monophosphorylated in unstimulated cells. Isoform in spot 60 kept the phosphorylation after stimulation and isoform 61 was not phosphorylated on this peptide at all. Phosphorylation on Thr189 seemed to be related to each specific isoform, and its regulation upon EGF stimulus appeared to be selective.

Phosphorylation on Thr189 behaved differently according to the isoform: it was present in Hsp5 isoform 59, not present at all in Hsp5 isoform 61, and present in a 1:1 (186-197P and 186-197hP; at 1313.6 m/z and 1319.6 m/z) in Hsp5 spot 60. Phosphorylation on Ser385 was appearing after stimulation for isoforms 59 and 60 but not for 61. It was not possible to calculate the extent of phosphorylation indirectly, because the isotopic patterns of peaks corresponding to 186-197 and 377-386hR overlapped.

Peptide 563-573 (1316.6 m/z) was present in all three spots, while 563-573PP (1476.6 m/z) was only present in spot 60. No conclusions could be drawn on the dynamics of phosphorylation upon stimulation because this peptide did not contain arginine, therefore these peptides are not included in the representation of the dynamics of phosphorylation for all Hsp5 isoforms identified (Figure 43).



### Figure 43. Phospho-kinetics of Hsp5

Distribution of phosphorylation upon stimulation with EGF in different Hsp5 isoforms. Each "worm" represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as -P. The name of the specific isoform analysed is indicated on the left.

### Protein that act as metabolic enzymes

Nine spots corresponding to three different proteins were metabolic enzymes (Table 6). For proteins Pyruvate kinase and S-adenosylhomocysteine hydrolase no PTMs could be identified from their MALDI-TOF spectra, despite the high sequence coverage. Glucose 6-Phosphate dehydrogenase was present in two different isoforms, and although no changes upon stimulation could be seen, only spot 126 showed phosphorylation on Tyr361 and Thr362, (it

can not be assessed if this is regulated by EGF signal because this peptide did not contain any arginine).

Spot	Protein that act as metabolic enzymes	Gi No	Refseq	Gene	cov.
112	S-adenosylhomocysteine hydrolase	8392878	NP_058897	Ahcy	30%
113	S-adenosylhomocysteine hydrolase	8392878	NP_058897	Ahcy	40%
117	S-adenosylhomocysteine hydrolase	8392878	NP_058897	Ahcy	21%
126	Glucose6-phosphate dehydrogenase	204197	AAA41179	G6PD	29%
147	Glucose6-phosphate dehydrogenase	204197	AAA41179	G6PD	29%
189	Pyruvate Kinase, M2 isozyme	1346398	P11981	PKM2	33%
190	Pyruvate Kinase, M2 isozyme	1346398	P11981	PKM2	45%
200	Pyruvate Kinase, M2 isozyme	1346398	P11981	PKM2	44%

**Table 6. Proteins that act as metabolic enzymes**

## Protein acting in cell motility

Proteins known to play a role in cells motility and structure are listed in Table 7. Gamma Actin, Septin2 and vimentin did not present any PTMs.

Spot	Proteins involved in cell motility or structure	Gi No	Refseq	Gene	cov.
89	Gamma-actin.	809561	CAA31455		24%
116	Septin2	16924010	NP_476489	Sep-02	19%
13	Vimentin	14389299	NP_112402.1	Vim	35%
14	Vimentin	14389299	NP_112402.1	Vim	41%

**Table 7. Proteins involved in cell motility or structure**

## Protein acting as signalling molecules

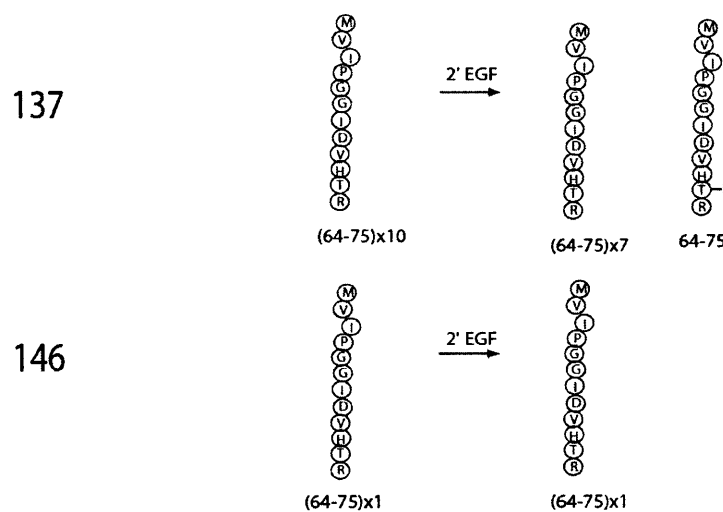
Four signalling proteins were identified (Table 8); of these RhoGDI-1 was present in two isoforms and showed one of the highest degrees of modification. The Signal Sequence Receptor delta does not contain any phosphorylation. Interestingly, RabGDI did not present any modifications either, suggesting a very different mechanism of regulation from RhoGDI.

spot	Proteins that act as signalling molecules	Gi No	Refseq	Gene	Cov.
137	Collapsin response mediator protein 2(CRMP-2)	1351260	P47942	Dpysl2	33%
146	Collapsin response mediator protein 2(CRMP-2)	1351260	P47942	Dpysl2	22%
221	RhoGDI-1	1346398	NP_001007006	Arhgdia	62%
222	RhoGDI-1	1346398	NP_001007006	Arhgdia	46%
224	Rab6, member RAS oncogene family	13195674	NP_077249	Rab6	45%
228	Rab1, RAS-related protein	45433570	NP_112352		57%
33	RabGDI beta	1707891	P50399	GDI3	49%
295	signal sequence receptor 4(delta)	8394364	NP_058895		30%

**Table 8. Proteins that act as signalling molecules**

### *Collapsin response mediator protein*

CRMP-2 was present in spots 137 and 146; only the isoform in 137 underwent phosphorylation on Thr74 after 2 minutes of stimulation with EGF (Figure 44). As for both isoforms the corresponding unphosphorylated peptide 64-75 was present in the spectra (peaks at 1294.7 m/z and 1300.7 m/z), the ratio of their corresponding peaks intensities before and after stimulation was calculated and the spectra normalised. In the isoform 137, the ratio of the unmodified peptide 64-74 from EGF treated cells to control cells was found to be 7:10. This indicated that 30% of this isoforms of CRMP-2 became phosphorylated on Thr74, 2 minutes after the cells have been stimulated with EGF. As expected, the ratio of the unmodified peptides 64-75 in isoform 146 was 1:1.



#### Figure 44. Phospho-kinetics of CRMP-2

Distribution phosphorylation upon stimulation with EGF in different CRMP-2 isoforms. Each “worm” represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as –P. The name of the specific isoform analysed is indicated on the left.

### *RhoGDI-1*

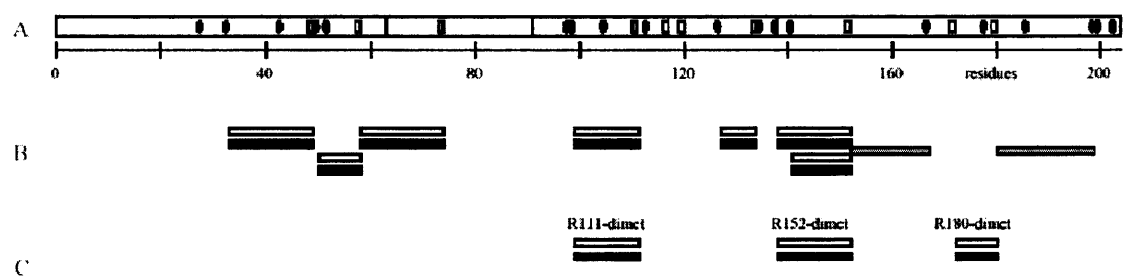
#### *RhoGDI-1: MALDI-TOF MS analysis*

RhoGDI-1 was present in two adjacent spots. The coverage obtained was above 60% for both isoforms. Arg117 and 172, which are expected to produce only very short tryptic peptides, were not observed in the present experiments. For peptides containing Arg111, 152 and 180, pairs of arginine-containing peptides showed masses that were unexpected on the basis of the amino acid sequence of RhoGDI-1, suggesting that these peptides carried additional post-translational modifications and some of these appeared to be other than phosphorylation. For this reason greater details are given on the analysis of both isoforms of RhoGDI. Comparison of the observed peptides with the exon structure of RhoGDI-1 showed that peptides were observed from all exons. This is good evidence that the full sequence of the protein was transcribed, even though only 66% of the sequence was covered by the tryptic peptides detected by MALDI-TOF MS (Figure 45). The experimental MW of about 30 kDa (theoretical 23.5 kDa) was also consistent with translation of the full protein.

The two isoforms share the majority of the doublets. The doublet of peaks 1601.7 m/z - 1607.7 m/z was assigned to the peptide 139-152 without any post-translational modifications. The doublet of peaks at 1629.8 m/z - 1635.8 m/z was assigned to dimethylated peptide 139-152. Peaks at 1454.8 m/z and 1460.8 m/z were assigned to peptide 100-111 and peaks at 1482.8 m/z and 1488.8 m/z were attributed to dimethylated peptide 100-111. Peaks at 952.5 m/z and 958.5 m/z ions were assigned to dimethylated 173-180 peptide. No evidence was found for this peptide without methylation. Peaks at 1565.5 m/z - 1571.6 m/z were attributed to peptide 142-152 with four phosphorylations. Such a negatively charged peptide is very difficult to ionise, and it was not possible to confirm its identity by ESI for this very same reason. Nevertheless it is mentioned amongst the possible modifications found. Specifically found only in isoform 221 is the doublet at 1670.7 m/z -1676.7 m/z, corresponding to peptide 100-111 with a dimethylation and a diphosphorylation.

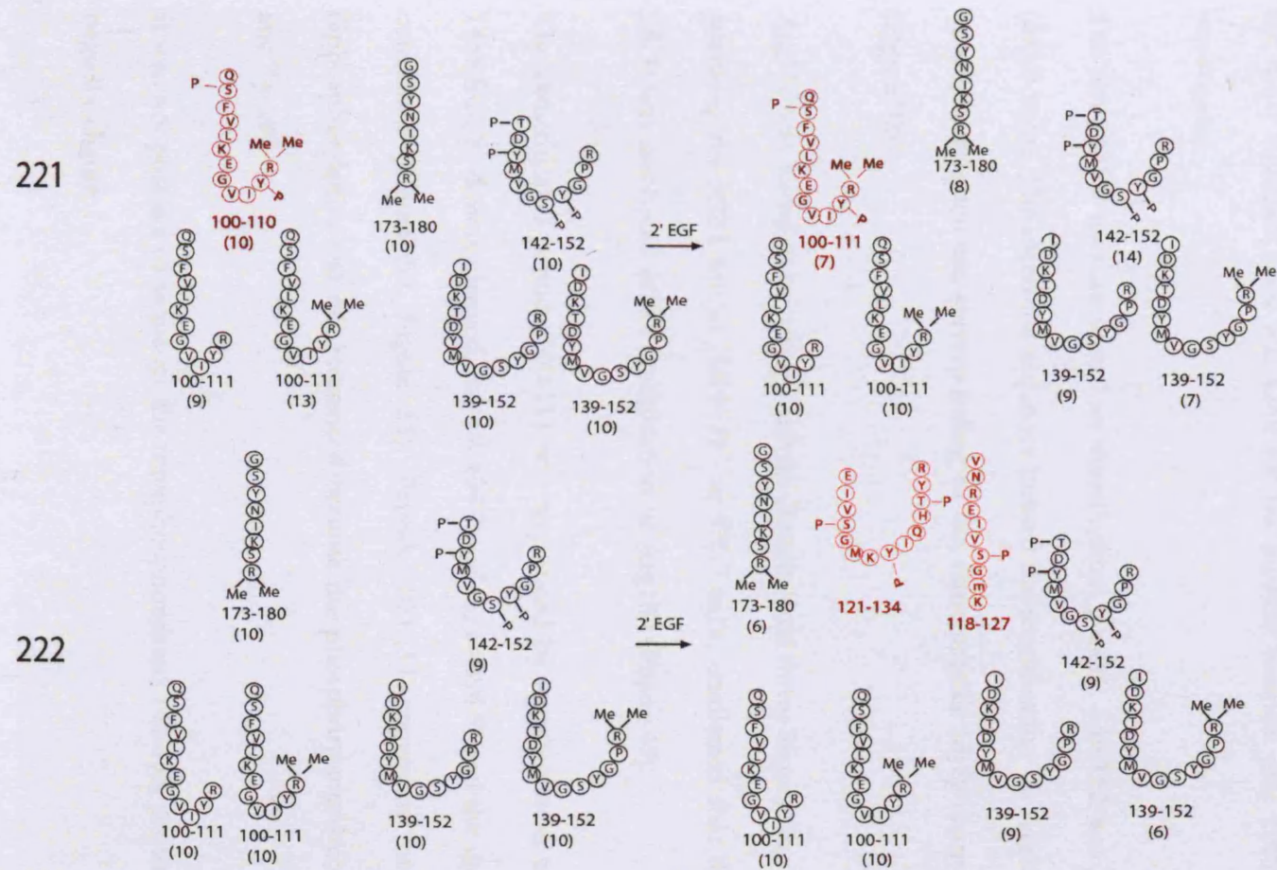
The isoform in spot 222 had two additional phosphopeptides specifically present only after EGF treatment. Despite the slightly lower coverage the following peaks were selectively present only in the spectrum of isoform 222: the peak at 1987.0 m/z corresponding to peptide 121-134 with three phosphorylations and the peak at 1234.6 m/z corresponding to monophosphorylated 118-127. Interestingly, these peptides overlap on the sequence of the protein, confirming the phosphorylation on Ser124. This is all summarised in Figure 46.

All peaks were confirmed to be part of doublets by comparison of the RhoGDI-1 spectrum from unlabelled cells.



**Figure 45. Summary of MS analyses of RhoGDI-1, spot 221**

(A) Exon boundaries (solid rectangles) and Arg (open bars) or Lys (solid ovals) residues. (B) Unmodified peptides identified by MALDI-TOF MS. Open / solid rectangles indicate peptides containing  $^{12}\text{C}$  /  $^{13}\text{C}$  arginine. Hatched rectangles indicate peptides cleaved at Lys. (C) Post-translationally modified peptides identified by MALDI-TOF MS and confirmed by MS/MS (see text).



**Figure 46. Phospho-kinetics of RhoGDI-1**

Distribution of phosphorylation upon stimulation with EGF in different RhoGDI-1 isoforms. Each “worm” represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as –P, while methylation is indicated by Me. The peptides highlighted in red correspond to the phosphopeptide specific for each isoform. The name of the specific isoform analysed is indicated on the left.



### *RhoGDI-1: ESI MS/MS analysis*

ESI MS/MS analyses were performed to confirm the identity of the peptides and to assign the post-translationally modified residues (Table 9). Only the data from isoform 221 are reported: the most abundant spot was used for the analysis because only common peaks have been investigated.

The ion 1635.8 m/z attributed to dimethylated peptide 139-152 was analysed as  $(MH+2)^{3+}$  (545.9 m/z). The identified sequence located the modification on Arg152; this was confirmed by sequencing the ion corresponding to the same peptide triply charged containing  $^{12}C_6$  Arg (Figure 48).

Arg180 was found to be present only in dimethylated form. Sequencing of the peptide 173-180 analysing the 952.5 ion as  $(MH+1)^{2+}$  at 476.7 m/z, confirmed that the increase in mass of 28Da was associated with a modification on Arg180 (Figure 49).

The dimethylated peptide 100-111 was sequenced by fragmentation of the ions 1482.8 and 1488.8 m/z, doubly charged, ( $m/z= 494.9$  and  $m/z=496.9$ ) and the dimethylation of Arg111 confirmed (Figure 50, Figure 51). Peptide 100-111 containing both dimethylation and diphosphorylation was not sequenced because the phosphorylations could only be on Ser101 and Tyr109.

It was not possible to sequence the tetraphosphorylated 142-152, probably because of its high negative charge.

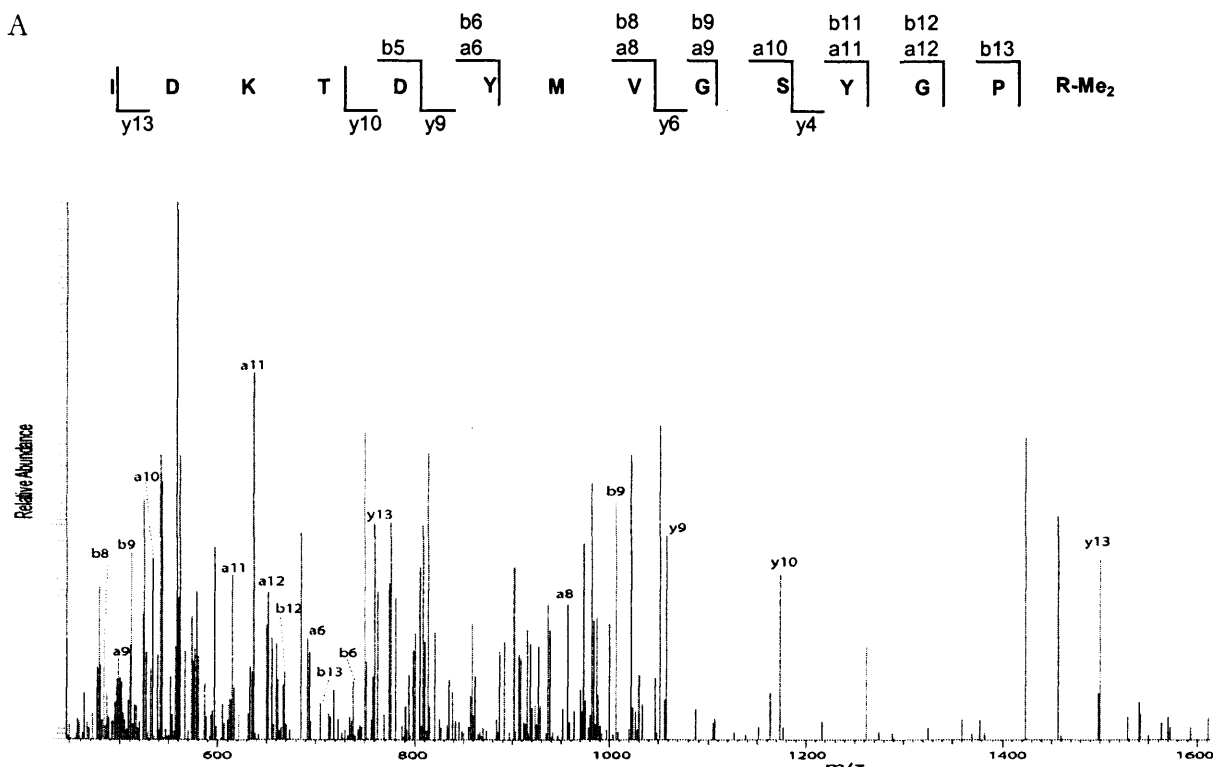
The phosphorylation sites on 121-134 of RhoGDI-1 in spot 222 could not be uniquely located.

Peptide	<sup>12</sup> C <sub>6</sub> Arg	<sup>13</sup> C <sub>6</sub> Arg	modifications	parent ion	charge	isolated ion	coverage
A 139-152	X		Arg dimethylation	1629.8	3+	543.9	I D K T D Y M V G S Y G P R-Me <sub>2</sub>
B 139-152		X	Arg dimethylation	1635.8	3+	545.9	I D K T D Y M V G S Y G P hR-Me <sub>2</sub>
C 173-180	X		Arg dimethylation	952.5	2+	476.7	G S Y N I K S R-Me <sub>2</sub>
D 100-111	X		Arg dimethylation	1482.8	3+	494.9	Q S F V L K E G V E Y R-Me <sub>2</sub>
E 100-111		X	Arg dimethylation	1488.8	3+	496.9	Q S F V L K E G V E Y hR-Me <sub>2</sub>

**Table 9. Summary of the peptides of interest and correspondent ions selected for ESI MS/MS analysis of RhoGDI-1**

The peptides analysed, the combination of modifications for each peptide, the molecular weight of the modified peptide, the charge of the ionised peptide, the corresponding m/z of the ion isolated and analysed, the coverage obtained are listed. For each peptide hR stands for <sup>13</sup>C labelled Arg; R-Me<sub>2</sub> and K-Me<sub>2</sub> stand for dimethylated arginine and lysine.

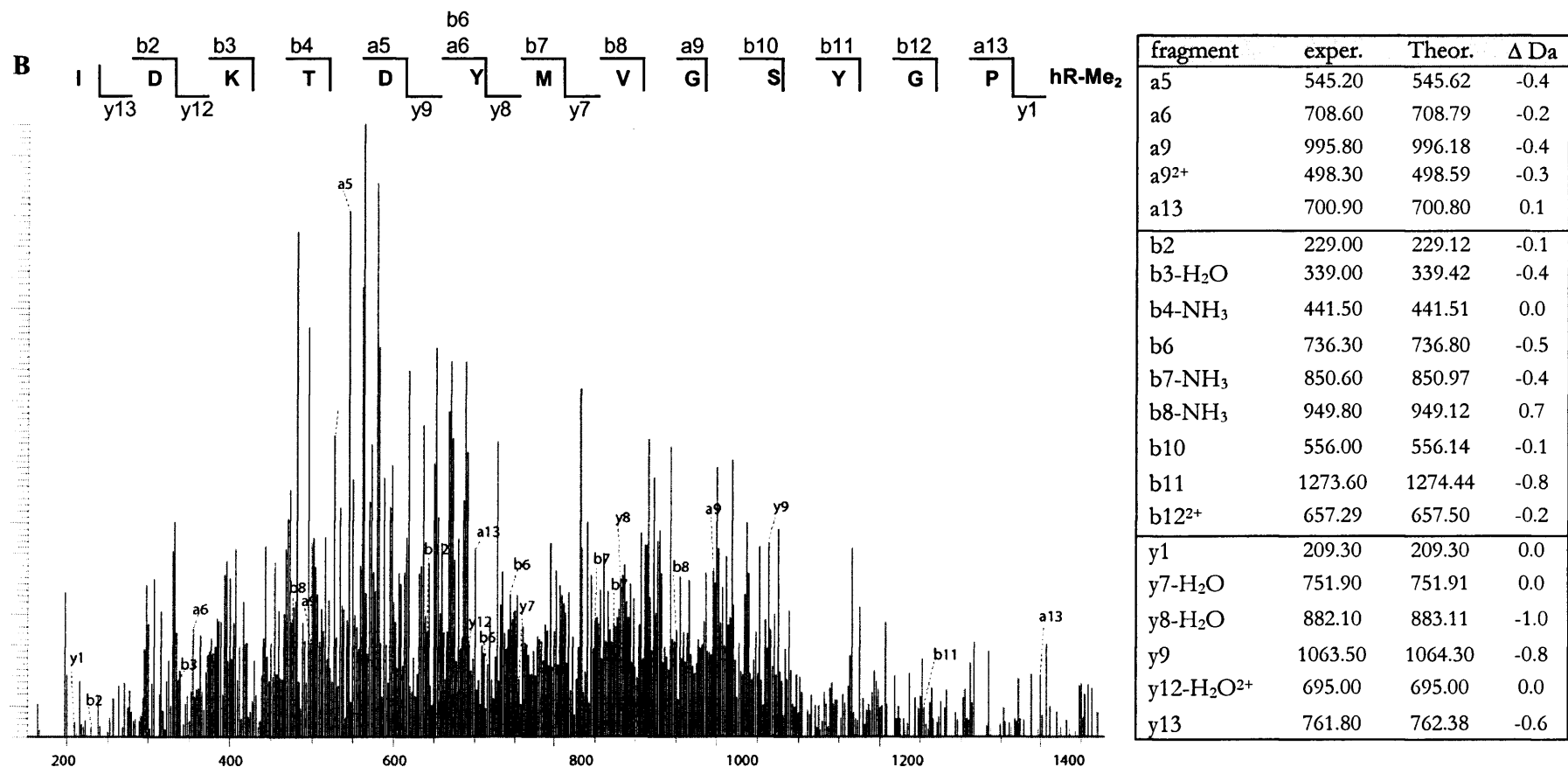
A



**Figure 47. ESI MS analysis of peptide A (as described in Table 9)**

The ions belonging to the peptide are annotated in the spectrum and they are listed in the table, together with the theoretical values and the errors associated, expressed as difference in Da. A schematic representation of the peptide with the corresponding ions identified is reported on the top. For convenience the charge and the NH<sub>3</sub>/H<sub>2</sub>O losses are not indicated in the spectrum and in the peptide scheme, but they can be found in the table. Me<sub>2</sub> is dimethylation.

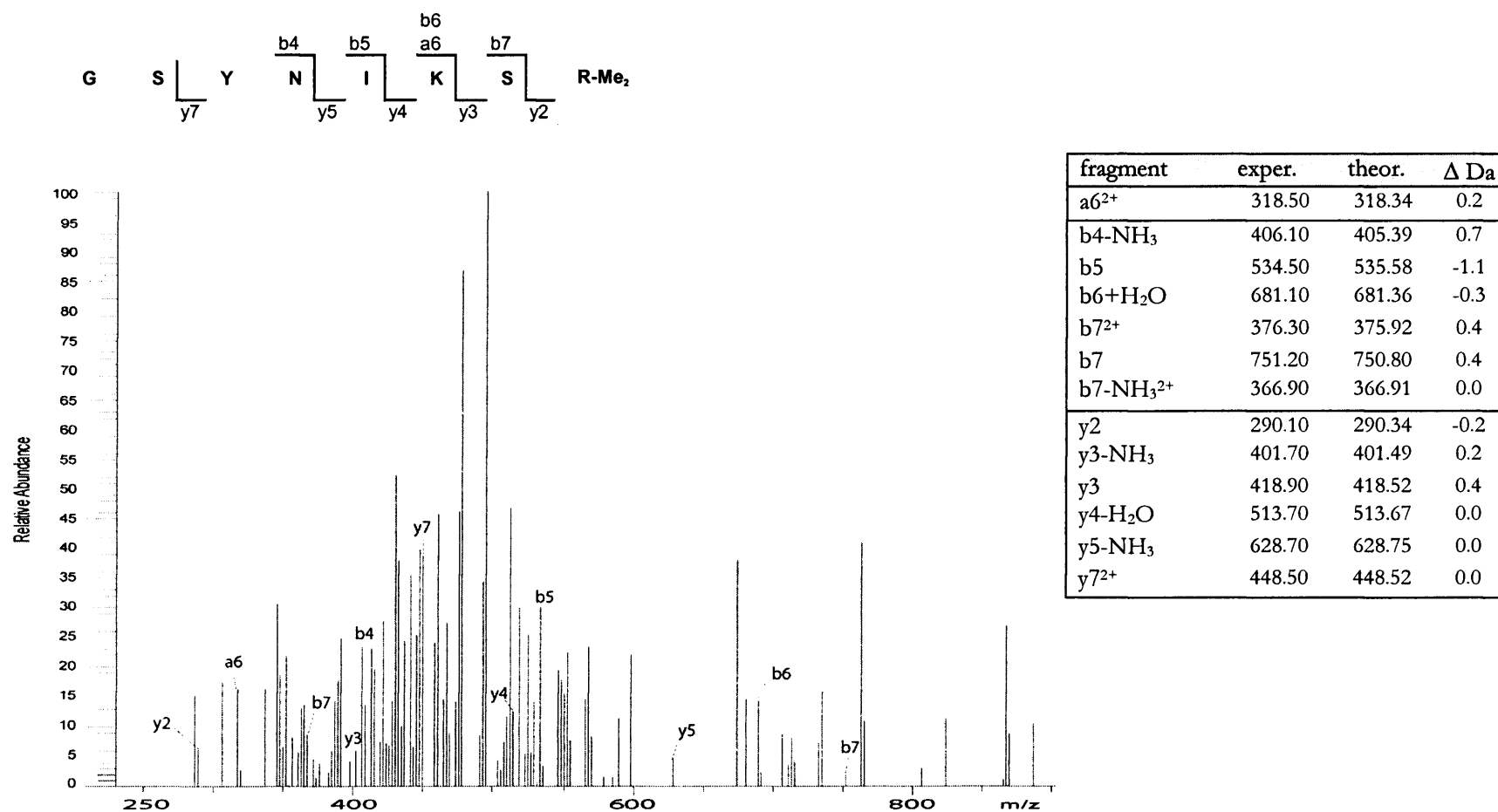
fragment	exper.	Theor.	$\Delta$ Da
a6-NH <sub>3</sub>	691.80	691.76	0.0
a8 <sup>2+</sup>	469.50	470.04	0.5
a8	938.10	939.13	1.0
a9 <sup>2+</sup>	499.00	498.59	-0.4
a10-NH <sub>3</sub> <sup>2+</sup>	532.90	533.62	0.7
a11-NH <sub>3</sub> <sup>2+</sup>	614.50	615.21	0.7
a12 <sup>2+</sup>	651.80	652.25	0.5
b5-H <sub>2</sub> O	556.80	556.60	-0.2
b6	736.80	736.80	0.0
b8 <sup>2+</sup>	483.30	484.07	0.8
b9-NH <sub>3</sub> <sup>2+</sup>	503.10	503.59	0.5
b9 <sup>2+</sup>	512.90	512.60	-0.3
b9	1006.20	1007.16	1.0
b11 <sup>2+</sup>	636.50	637.73	1.2
b12 <sup>2+</sup>	665.90	666.25	0.4
b12-H <sub>2</sub> O <sup>2+</sup>	704.90	705.80	0.9
b13 <sup>2+</sup>	714.60	714.81	0.2
y4	520.80	520.61	-0.2
y6	664.70	664.74	0.0
y9	1057.50	1058.25	0.8
y10	1173.30	1173.34	0.0
y13 <sup>2+</sup>	758.40	758.86	0.5
y13-NH <sub>3</sub>	1499.40	1499.70	0.3



**Figure 48. ESI MS analysis of peptide B (as described in Table 9)**

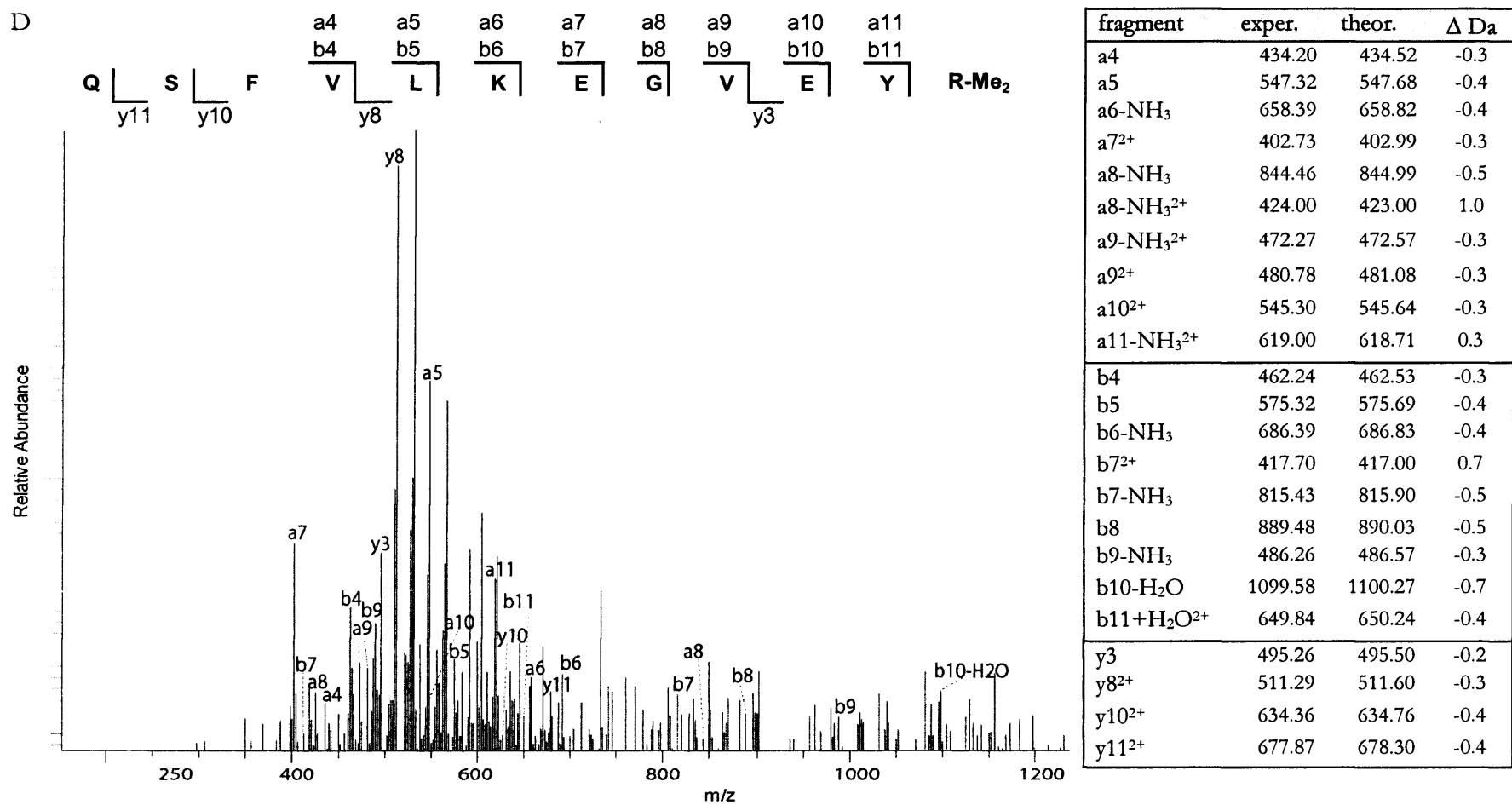
The ions belonging to the peptide are annotated in the spectrum and they are listed in the table, together with the theoretical values and the errors associated, expressed as difference in Da. A schematic representation of the peptide with the corresponding ions identified is reported on the top. For convenience the charge and the NH<sub>3</sub>/H<sub>2</sub>O losses are not indicated in the spectrum and in the peptide scheme, but they can be found in the table. Me<sub>2</sub> is dimethylation, hR is <sup>13</sup>C labelled Arg.

C



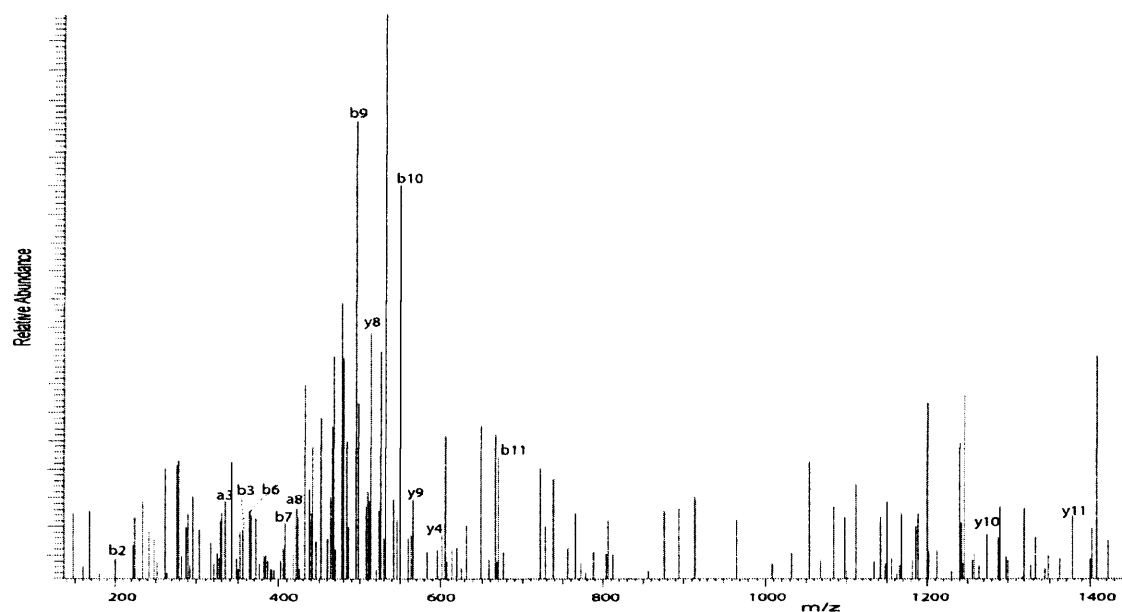
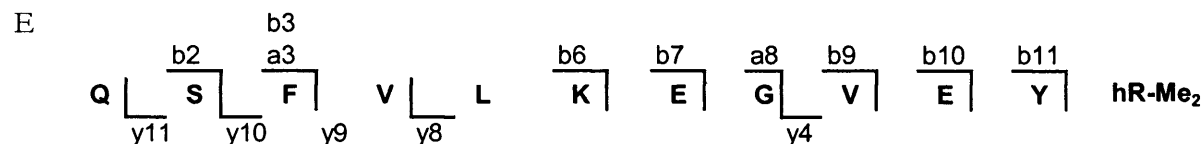
**Figure 49. ESI MS analysis of peptide C (as described in Table 9)**

The ions belonging to the peptide are annotated in the spectrum and they are listed in the table, together with the theoretical values and the errors associated, expressed as difference in Da. A schematic representation of the peptide with the corresponding ions identified is reported on the top. For convenience the charge and the NH<sub>3</sub>/H<sub>2</sub>O losses are not indicated in the spectrum and in the peptide scheme, but they can be found in the table. Me<sub>2</sub> is dimethylation.



**Figure 50. ESI MS analysis of peptide D (as described in Table 9)**

The ions belonging to the peptide are annotated in the spectrum and they are listed in the table, together with the theoretical values and the errors associated, expressed as difference in Da. A schematic representation of the peptide with the corresponding ions identified is reported on the top. For convenience the charge and the NH<sub>3</sub>/H<sub>2</sub>O losses are not indicated in the spectrum and in the peptide scheme, but they can be found in the table. Me<sub>2</sub> is dimethylation.



fragment	exper.	theor.	$\Delta$ Da
a3	334.90	335.39	-0.5
a8-NH <sub>3</sub> <sup>2+</sup>	422.80	423.00	-0.2
b2-NH <sub>3</sub>	199.20	199.19	0.0
b3	363.20	363.40	-0.2
b6 <sup>2+</sup>	352.50	352.44	0.1
b7-NH <sub>3</sub> <sup>2+</sup>	408.10	408.48	-0.4
b9 <sup>2+</sup>	496.40	495.09	1.3
b10-H <sub>2</sub> O <sup>2+</sup>	549.90	550.64	-0.7
b11	671.70	672.81	-1.1
y4	600.03	600.74	-0.7
y8 <sup>2+</sup>	514.30	514.63	-0.3
y9 <sup>2+</sup>	564.80	564.19	0.6
y10	1273.90	1274.56	-0.7
y11	1361.30	1361.63	-0.3

**Figure 51. ESI MS analysis of peptide E (as described in Table 9)**

The ions belonging to the peptide are annotated in the spectrum and they are listed in the table, together with the theoretical values and the errors associated, expressed as difference in Da. A schematic representation of the peptide with the corresponding ions identified is reported on the top. For convenience, the charge and the NH<sub>3</sub>/H<sub>2</sub>O losses are not indicated in the spectrum and in the peptide scheme, but they can be found in the table. Me<sub>2</sub> is dimethylation, hR is <sup>13</sup>C labelled Arg.

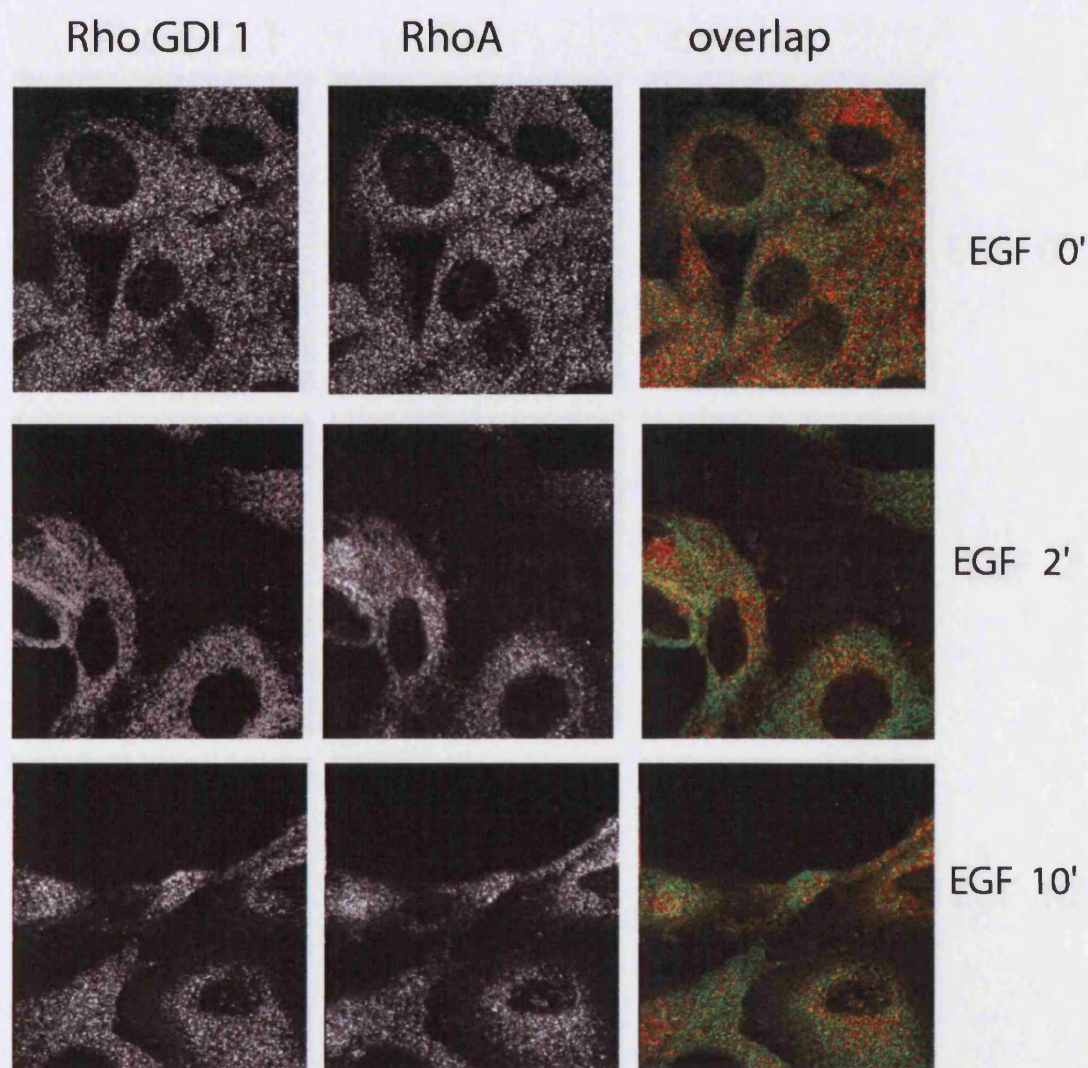
### *RhoGDI-1: Immunofluorescence study*

In an attempt to clarify the role of the modifications in RhoGDI, immunofluorescence studies were performed to ascertain if PTMs would alter cellular location and the degree of overlap of GDI with Rac and RhoA, indicative of the binding.

The cytosolic distribution of RhoGDI-1 has been widely observed before (Gibson et al. 2001; Koch et al. 1997; Shisheva et al. 1994), however no immunofluorescence has been performed early on in EGF stimulation.

Interestingly RhoA and Rac showed a different distribution within the cell, and this is independent from EGF stimulation. Rac seemed to be concentrated around the cell membrane and can be seen in ruffles in accordance to published work (Topp et al. 2004) (Figure 53). RhoA was spread equally in the cytosol. However, no apparent changes in GDI location could be detected (Figure 52).

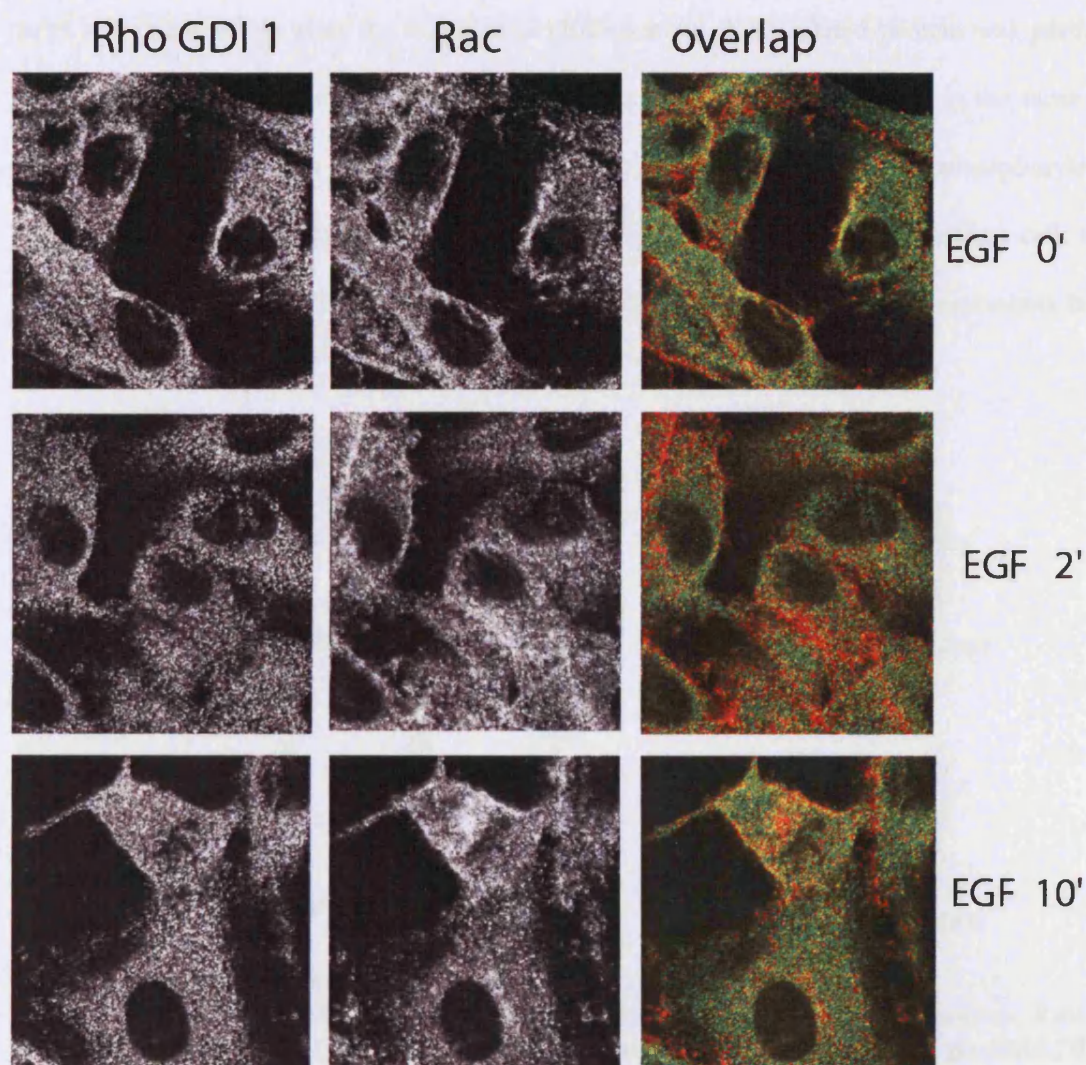




**Figure 52. RhoGDI-RhoA immunofluorescence**

The first column represents the green channel, corresponding to the staining with RhoGDI-1 antibody, the second column is the red channel, corresponding to RhoA antibody staining, and the third column is an overlap of the two. Each row corresponds to cells stimulated for the indicated times.



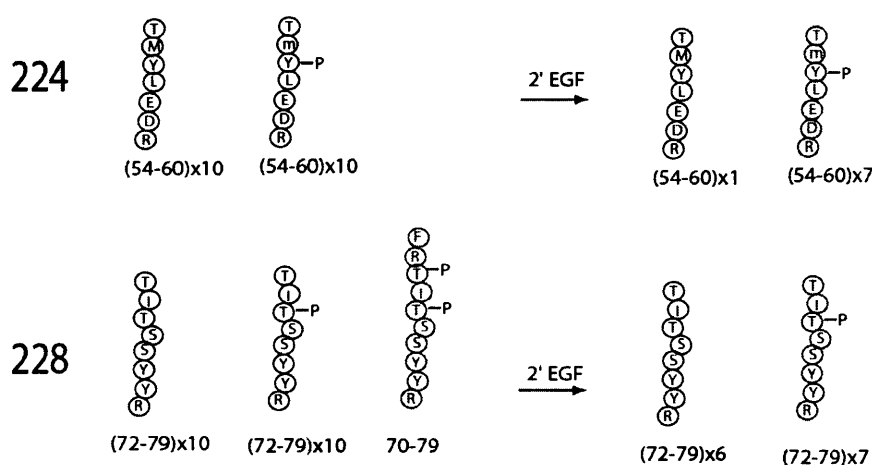


**Figure 53. RhoGDI-Rac immunofluorescence**

The first column represents the green channel, corresponding to the staining with RhoGDI-1 antibody, the second column is the red channel, corresponding to Rac antibody staining, and the third column is an overlap of the two. Each row corresponds to cells stimulated for the indicated times.

## Rab6 and Rab1

Rab6 and RAS-related proteins (similar to Rab1) showed phosphorylation. Rab6 was already partially phosphorylated on peptide 54-60 in quiescent cells, (corresponding peak at 1023.6 m/z) and remained so after the stimulation (1029.6 m/z). RAS-related protein was partially phosphorylated before and after the stimulation in a similar way to Rab6 and in the same N-terminal region. A doublet at 1070.5 m/z -1076.5 m/z corresponding to monophosphorylated 72-79 was found. An additional diphosphorylated peptide present only in quiescent cells was also identified (1453.7 m/z corresponding to 70-79). The following scheme represents both proteins (Figure 54).



**Figure 54. Phospho-kinetics of Rab1 and Rab6**

Distribution of phosphorylation upon stimulation with EGF in different Rab proteins, Rab6 (224) and similar to Rab1 (228). Each "worm" represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as -P. The name of the specific isoform analysed is indicated on the left.

## Proteins involved in protein synthesis

Seven different proteins involved in protein synthesis were found (Table 10). For most of those no phosphorylations could be identified. For the protein contained in spot 87 only one monophosphorylated peptide was recognised (369-381, at 1513.7 m/z), but it did not contain

any arginine. It can only be suggested that this protein is phosphorylated on Thr377, but no connections of this event to the EGF signal can be made.

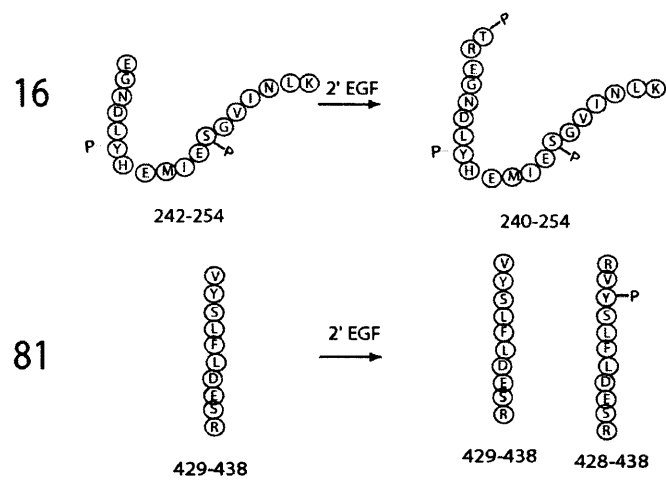
spot	Proteins involved in protein synthesis	Gi No	Refseq	Gene	Cov.
	Similar to Eukaryotic initiation factor				
87	4A-I	40786436	NP_955404		50%
151	Aspartyl-tRNA synthetase	16758642	NP_446251	Dars	13%
	ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, beta subunit				
16		54792127	NP_599191	Atp5b	36%
243	Acidic ribosomal protein P0	11693176	NP_071797	Arbp	50%
44	Elongation factor 2	119176	P05197	EEF2	18%
				LOC2	
119	Similar to Elongation factor 1-gamma	34861590	XP_215165	93725	25%
81	RuvB-like protein 2	6755382	NP_035434	Ruvbl2	24%

**Table 10. Proteins involved in protein synthesis**

### *ATP synthase beta and RuvB-like protein 2*

Interesting behaviour was observed for ATP synthase beta. One phosphopeptide was identified: peptide 242-254 is present diphosphorylated (at 2221.1 m/z) but it does not contain arginine, therefore no conclusion on the kinetic of these phosphorylations can be made. A third phosphorylation in response to EGF stimulus was found (peptide 240-254, at 2564.0 m/z), and could be associated with the missed cleavage on Arg241.

RuvB-like protein 2 showed a single phosphorylation upon stimulation on 428-438 (428-438hR-P peak at 1476.7 m/z). The missed cleavage on Arg428 suggests that the phosphorylation might be close to it (maybe Tyr430).



**Figure 55. Phospho-kinetics of synthesis related proteins**

Distribution of phosphorylation upon stimulation with EGF of ATP synthase beta (16) and RuvB-like protein 2 (81). Each “worm” represents a peptide observed in the MALDI-TOF spectrum of the corresponding protein. Phosphorylation on different residues is indicated as -P. The name of the specific isoform analysed is indicated on the left.

## **DISCUSSION**

### **Proteins involved in EGF signalling, not recognised by kinase substrate antibodies**

The Phospho-SILAC method was applied to a wide range of proteins in order to investigate common behaviours in different classes of proteins upon stimulation with EGF.

Three enzymes, each present in 2-3 isoforms, and three structural proteins with up to two isoforms were identified, but despite the high coverage obtained no modifications could be found. All these proteins have been extracted using the phosphoprotein enrichment protocol (chapter 3), so they are likely to have at least a certain level of basal phosphorylation. However, no phosphopeptides were seen, and no unmodified peptides deviated from the 1:1 ratio, suggesting that none of the peptides identified could be a potential carrier. The phosphorylation must lie in the part of the protein which was not detected by MALDI-TOF MS. This seems particularly true for proteins like actin, which is known to undergo rearrangement and phosphorylation. Another point to be made is that these proteins are very abundant and therefore very low levels of phosphorylation (below 10%) would not be detectable.

A range of interesting data was obtained for the subset of signalling proteins identified; this is due not only to the fact that these proteins are expected to be involved in the signalling through their phosphorylation, but also because they are generally smaller, and a larger coverage could be obtained.

### ***Different chaperones, same phosphorylation site***

The family of chaperones has proved very interesting and has been extensively discussed in previous chapters. A total of 13 analysed spots contained chaperones (including Hsp8), two

were found to be Hsp4 and three Hsp5. In short, Hsp4 and Hsp5 underwent phosphorylation changes upon EGF treatment, but not in sites recognised by the antibody used in the initial screening of the NRK49F phosphoproteome (chapter 4). Other chaperones did not present any changes in the first two minutes of stimulation.

All these proteins belong to the Hsc70/Hsp70 family, and have been aligned so that their level of similarity could be visualised (Figure 56). The Thr38, found to be phosphorylated in Hsp8, is conserved in the three proteins (together with the P in +1, creating the consensus for CDK/MAPK). However, Thr38 has been found to be phosphorylated only in Hsp8, which explains why only this protein is recognised by the antibody against phosphorylated substrates of MAPK. Hsp8 is also phosphorylated on Thr163, which is aligned with Thr189 in Hsp5. These two residues have both been found to be phosphorylated in this study, although with different dynamics. In addition, it needs to be stressed that the close surrounding of these proteins is identical, but the whole tryptic peptide is not, leading to two different masses in the two spectra analysed and giving more relevance to the attribution of this phosphorylation.

Another general conclusion is that all proteins identified belonging to the family of Hsc70/Hsp70 present differential phosphorylation in the N-terminal of the ATPase domain, suggesting that the regulation of the activity happens primarily in this domain.



Hsp8	-----MSKGPVAGIDLGTTYSCVGVFQHGKV	26
Hsp5	MKFTVVAAALLLLCAVRAEEEDKKEDVGTVVGIDLGTTYSCVGVFKNGRV	50
Hsp4	-----MSVVGIDLGFGQSCYVAVARAGGI	23
	..***** .*. : * :	
Hsp8	EIIANDQGNRTTPSYVAFT-DTERLIGDAAKNQVAMNPTNTVFDARLIG	75
Hsp5	EIIANDQGNRITPSYVAFTPEGERLIGDAAKNQLTSNPENTVFDARLIG	100
Hsp4	ETIANEYSDRCTPACVSFG-PKNRSVGAAAKSQVISNAKNTVQGFKRFHG	72
	* ** : . : * ** : * : * : * : * : * : * : * : *	
Hsp8	RRFDDAVVQSDMKHWPFMVNDAGRPKVQVEYKG-ETKSFYPEEVSSMVL	124
Hsp5	RTWNDPSVQQDIKFLPFKVVEKTKPYIQVDIGGGQTKTFAPEEISAMVL	150
Hsp4	RAFSDPFVEAEKSNLAYDIVQLPTGLTGIKVTYMEEERNFTTEQVTAMLL	122
	* : . * . * : : . : : * : : : * : * : * : * : *	
Hsp8	TKMKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKDAGTIAGLNVLRIN	174
Hsp5	TKMKETAAYLGKKVTHAVVTVPAYFNDAQRQATKDAGTIAGLNVMRIIN	200
Hsp4	SKLKETAESVLKKPVVDCVSVSPSYTDAERRSVM DATQIAGLNCRLMN	172
	: * : * * : * * * : * : * : * : * : * : * : * : *	
Hsp8	EPTAAAIAYGLDK-----KVGAEARNVLIFDLGGGTFDVSILTIEDGIFEV	219
Hsp5	EPTAAAIAYGLDK-----REG-EKNILVFDLGGGTFDVSLLTIDNGVFEV	244
Hsp4	ETTAVALAYGIYKQDLPALEEKPRNVFVDMGHSAYQVSVCAFNRGKLKV	222
	* . * . * : * : * : * : * : * : * : * : * : * : *	
Hsp8	KSTAGDTHLGGEDFDNRMVNHFAIEFKRKHKKDISENKRAVRRRLRTACER	269
Hsp5	VATNGDTHLGGEDFDQVMEHFIKLYKKKTGKDV RKDNRAVQKLRREVEK	294
Hsp4	LATAFDTTLGGRKFDEVLVNHFCFEFGKKYKLDIKSKVRALLRLSQECEK	272
	: * * * * * : * : * : * : * : * : * : * : * : *	
Hsp8	AKRTLSS-STQASIEIDSLYEGIDFYTSITRARFEELNADLFRGTLDPVE	318
Hsp5	AKRALSS-QHQARIEIESFFEGEDFSETLTRAKFEELNMDLFRSTMKPVQ	343
Hsp4	LKKLMSANASDLPLSIECFMNDIDVSGTMNRGKFLEMCDLLARVEPLR	322
	* : : : : : * : : * : * : * : * : * : * : * : *	
Hsp8	KALRDAKLKDSQIHDIIVLGGSTRIPKIQKLLQDFFNGKELNKSINPDEA	368
Hsp5	KVLEDSLKKSIDIIVLGGSTRIPKIQQLVKEFFNGKEPSRGINPDEA	393
Hsp4	SILDQSKLKKEDIYAVEIVGGATRIPAVKEKISKFFG-KELSTTLNADEA	371
	. * : : * . * : * : * : * : * : * : * : * : * : *	
Hsp8	VAYGAAVQAAILSGDKSENVQDLLLLDVTPSLGIETAGGVMTVLIKRNT	418
Hsp5	VAYGAAVQAGVLSGD--QDTGDLVLLDVCPLTLGIETVGGVMTKLIPRNT	441
Hsp4	VTRGCALQCAILSPAFAKREFSITDVVPYPISLRWNSPAEEGSSDCEVFP	421
	* : * . * : * : * : * : * : * : * : * : * : *	
Hsp8	TIPTKQTQFTTYSNQPVGVLIVYEGERAMTKDNNLLGKFELTGIPPAP	468
Hsp5	VVPTKKSQIFSTASDNQPTVTIKVYGERPLTKDNHLLGTFDLTGIPPAP	491
Hsp4	KNHAAPFSKVLTFYRKEPFTLEAYYSSPDLPYPDPAIAQFSVQKVTPQS	471
	: . . * : * . * . : : : : * : : * : *	
Hsp8	RG-VPQIEVTFDIDANGILNVS-----AVDKSTGKENK	500
Hsp5	RG-VPQIEVTFEIDVNGILRVT-----AEDKGTGNKNK	523
Hsp4	DGSSSKVKVKVRNVHGFVSSAALVEVHKSESESEPMETDQNAKEEEK	521
	* : : : * . : : * : * : * : * : * : * : * : *	
Hsp8	ITITNDK-----	507
Hsp5	ITITNDQ-----	530
Hsp4	MQVDQEEPHTEEQQPQTPAENKAEESEMETSQAGSKDKKMDQPPQAKKAK	571



	: : ::	
Hsp8	-----GRLSKEDIERMVQEAKEYKADEKQDKVSSKNSL	542
Hsp5	-----NRLTPEEIERMVNDAEKFAEEDKKLKERIDTRNEL	565
Hsp4	VKTSTVDLPIESQLLWQLDREMLGLYTENEGKMIMQDKLEKERNDKNAV	621
	: * * : . : * : * : : : * :	
Hsp8	ESYAFNMKATVED-----	555
Hsp5	ESYAYSLKNQIGDK-----	579
Hsp4	EEYVYEMRDKLSGEYEKFVSEDDRNFTLKLEDTENWLYEDGEDQPKQVY	671
	* . * . . . : .	
Hsp8	-----EK	557
Hsp5	-----EK	581
Hsp4	VDKLAELRTLQPIKTRFQSEERPKLFEELGKQIQQYMKVISSFKNKED	721
	* .	
Hsp8	LQGKINDEDKQKILDKCNEIISWLDKNQTAE-----KEEFEHQ	595
Hsp5	LGGKLSPEDKETMEKAVEEKIEWLESHQDAD-----IEDFKAK	619
Hsp4	QYEHLDAAADMTKVEKSTNEAMEWMNSKLNQNKQSLTADPVVKTKEIEAK	771
	: : . * . : . : * : . * : : : : : : : :	
Hsp8	QKELEKVCNPIITKLYQ-----SAGGMPGGMPGGFP	626
Hsp5	KKELEEIVQPIISKLYG-----SGG-----	639
Hsp4	IKELTNICSPIISKPKPKVEPPKEEPKHAEQNGPVDGQGDNPQTQAAEHG	821
	*** : : . *** : *	
Hsp8	GGGAPPSGGASSGPTIEEVD	646
Hsp5	----PPPTGEEDTSEKDEL-	654
Hsp4	ADTAVPSDGDKKLPEMDID-	840
	* . * . . . :	

**Figure 56. Protein sequence alignment of Hsp8, Hsp5 and Hsp4**

Residues in red have been found to be phosphorylated, and the phospho-residues common to different isoforms were further highlighted in yellow. In green, the ATPase domain; in blue the sequence recognised by TPR-containing cochaperones. Alignment obtained using Clustal W 1.82 with default settings.

### *Proteins involved in protein synthesis*

ATP synthase beta is the catalytic subunit of ATP synthase producing ATP from ADP in presence of a proton gradient across the mitochondrial membrane (Vendemiale et al. 1995). No connection at the level of PTMs has yet been made with EGF signalling, although some observations relate long treatment with nafenopin to decreased the expression of ATP synthase beta (Chevalier et al. 2000). The general architecture of the enzyme is conserved among species; in E.Coli this consists of a globular catalytic moiety F(1) (corresponding to beta), protruding out of the inner side of the membrane, a membrane integral proton translocating moiety F(o) (alpha), and a stalk connecting F(1) to F(0) (Gaballo et al. 2002).

The most closely related structure is 1mab (rat liver F1-ATPase). The amino acids of interest are all conserved. All the residues found to be phosphorylated lie on the surface of the protein. In particular Tyr247 establishes intra-molecular interactions, but is also close to the subunit alpha and could easily host a phosphate moiety. Ser253 is exposed, but very distant from the membrane domain. The residue that can most likely influence the binding with the alpha subunit is Thr240 when phosphorylated.

Ruv2-like protein is a DNA Helicase, and although much is known about it, no clear connections with EGF signalling have been made. However, phosphorylation has been observed in the xeroderma pigmentosum group B (XPB) helicase, involved in nucleotide excision repair (NER) and transcription. The Ser751 residue of XPB was found to be phosphorylated in vivo and this phosphorylation was shown to inhibit DNA repair (Coin et al. 2004). Structures are available for the catalytic domains of the RuvB2-like protein but not for the C-terminal region where a potential phosphorylation on Tyr430 is placed.

### *Signalling proteins*

Seven spots were identified as signalling proteins, namely Collapsin Response Mediator Protein 2, two members of Rab protein family, RabGDI beta and RhoGDI-1.

Collapsin response mediator protein (CRMP-2) which also has been independently identified as Ulip2/CRMP-62/TOAD-64/DRP-2, is one of at least five isoforms (Arimura et al. 2004; Goshima et al. 1995). CRMP-2 is known to be expressed exclusively and highly in the developing nervous system and shows the ability to convert immature neurites and preexisting dendrites to axons. A discussion with Dr. Goshima has highlighted the possibility that this protein is present in lung epithelial cells (Ito et al. 2000; Kagoshima et al. 2001). His recent observations suggest that CRMP-2 has roles in cell polarity as well as in mediating external

signals could play a role in fibroblasts as well. A large complex including atypical protein kinase C (aPKC) complex accumulates at the tip of the axon, and its polarised localisation and aPKC activity are important for axon specification (Shi et al. 2003). aPKC can also phosphorylate GSK-3 $\beta$  and inactivate its kinase activity, and GSK-3 $\beta$  is important for polarisation of migrating fibroblasts (Etienne-Manneville et al. 2003). CRMP-2 is phosphorylated in the brain by Rho-kinase on Thr555 (Arimura et al. 2004) and phosphorylation on Thr509, Ser522 and Ser518 has also been reported in relation with Alzheimer disease (Gu et al. 2000). However, in fibroblasts CRMP-2 is not phosphorylated on Ser522 upon stimulation with EGF (data not shown). Very recently, it has been shown that GSK-3 $\beta$  participates in neuronal polarisation through CRMP-2 by phosphorylation of CRMP-2 at Thr514. NT-3 and brain-derived neurotrophic factor (BDNF) inhibit GSK-3 $\beta$  via the phosphatidylinositol-3-kinase (PI3-kinase)/PKB pathway, thereby reducing phosphorylation levels of CRMP-2 at Thr514, leading to axon elongation and branching (Yoshimura et al. 2005). In a recent review a model has been built that inserts CRMP-2 as a crucial signal protein in the growth factor signalling for axon development.

### *RhoGDI-1*

Arginine methylation of RhoGDI-1 has not been previously reported. The present results suggest that RhoGDI-1 is partially dimethylated on Arg111 and 152 and fully dimethylated on Arg180. In fact, for Arg111 and Arg152 both unmodified and dimethylated forms were observed. For Arg180, the boundaries of the tryptic peptides were such that definite evidence for the dimethylated forms was obtained, but not for the unmodified residues. Comparison of the labelled and unlabelled peptides shows that the ratio for methylated peptides fluctuates around the value of 1:1, independently from EGF treatment. This indicates that stimulation

with EGF has no substantial effect on dimethylation of the arginine residues 111, 152 and 180 at short times.

For both isoforms multiple phosphorylations on the peptide 142-152 was observed equally before and after stimulation with a ratio close to 1:1.

However, only isoform 222 presented phosphorylation on Ser124 and two additional phosphorylations on two of the following residues: Tyr128, Thr132 and Tyr133. At the same time isoform 222 showed diphosphorylation on methylated 100-111, on Ser101 and Tyr110, which seemed to decrease slightly upon stimulation.

The only other residues known to be phosphorylated in RhoGDI-1 are Ser101 and Ser174. Very recently, it has been shown that phosphorylation of both Ser101 and Ser174 by the kinase Pak-1 leads to dissociation of Rac/RhoGDI-1 complexes, but not of RhoA/RhoGDI-1 complexes (DerMardirossian et al. 2004). Ser101 and 174 are both located away from the Rac/RhoGDI-1 interaction surface and can apparently be phosphorylated in the complex, so that the mechanism of dissociation is not entirely clear (DerMardirossian et al. 2004). This phosphorylation will be isoform-specific since the two serines are changed to threonine in RhoGDI-2 and to Val and Pro in RhoGDI-3. Phosphorylation of Ser101/174 by Pak-1 and dissociation of Rac/RhoGDI-1 complexes was reported following stimulation of HeLa cells with EGF (DerMardirossian et al. 2004). In the present experiments only phosphorylation on Ser101 was observed, although with different dynamics, and in the presence of dimethylation and an additional phosphorylation on Tyr110. This suggests that for EGF stimulation, downstream pathways involving RhoGDI-1 may be dependent on the cell type.

A crude examination was performed to understand if the PTMs on RhoGDI-1 could result in a change in cellular location upon stimulation. Both GDI and GTPases are widely present in the

cytosol and no changes could be visualised upon stimulation. This is probably due to the fact that only a small fraction of GTPases and RhoGDI-1 is involved at any time in signalling.

The post-translational modifications observed are in agreement with the known structure and functional roles of RhoGDI-1. The structure of RhoGDI-1 consists of a C-terminal immunoglobulin-like domain with a pocket that binds the geranyl-geranyl modification at the C-terminus and a flexible N-terminus which becomes helical on interaction with Rho family proteins (Gosser et al. 1997; Keep et al. 1997). The structures of four complexes of RhoGDI-1 with Rho family proteins have been solved by X-ray crystallography. These vary in how well the flexible 59 residue N-terminus of RhoGDI-1 and the C-terminal basic and isoprenylated 10 residue region of the G-protein are resolved. RhoGDI-1 with Rac1 (PDB 1HH4) (Grizot et al. 2001) has two copies of the complex, in both cases some density for the geranyl-geranyl group is seen, but it is not fully connected to the C-terminus of the Rho protein. Cdc42 (PDB 1doa) (Hoffman et al. 2000) has the best defined C-terminus of the G-protein with full connectivity for the C-terminal methylated cysteine and the geranyl-geranyl group. Both these structures have more or less the full RhoGDI-1 resolved. A low resolution structure with RhoA (PDB 1cc0) (Longenecker et al. 1999) shows neither the N-terminus of RhoGDI-1 nor the C-terminal part of RhoA. The structure of Rac-2 with RhoGDI-2 (PDB 1ds6) (Scheffzek et al. 2000) shows some of the N-terminus of RhoGDI-1 but did not resolve the isoprenylation region of the G-protein.

The sites of the present modifications can be mapped onto the above mentioned structures. Ser148 lies in the interface between the G-protein and the GDI, where it forms a hydrogen bond to a side chain nitrogen of His104 of the Rac or Cdc42 G-proteins. Phosphorylation would disrupt this interaction. In RhoA this histidine is not conserved, suggesting that phosphorylation of Ser148 may help RhoGDI-1 to discriminate amongst Rho family members.

Tyr144 hydrogen bonds to Arg186 of Cdc42 two residues away from the geranyl-geranylated C-terminal cysteine. This area of the Rho family of G-proteins is rich in basic residues and the most variable between family members. The phosphorylation sites thus have the potential to alter the stability of the G-protein/GDI complex, potentially in an isoform-specific manner. In the two different copies in the Rac1/RhoGDI-1 crystal structures, different basic residues Lys188 and Arg187 (1 and 2 residues from the C-terminus respectively) are close to Tyr144, but further than hydrogen bonding distance. This area is unresolved in the other two structures. The effect of phosphorylation could potentially either increase the affinity for the basic residues or decrease the affinity by disrupting hydrogen bonding interaction to the same basic residues. Tyr144 is a phenylalanine in RhoGDI-2, therefore this phosphorylation may discriminate amongst RhoGDI-1 isoforms.

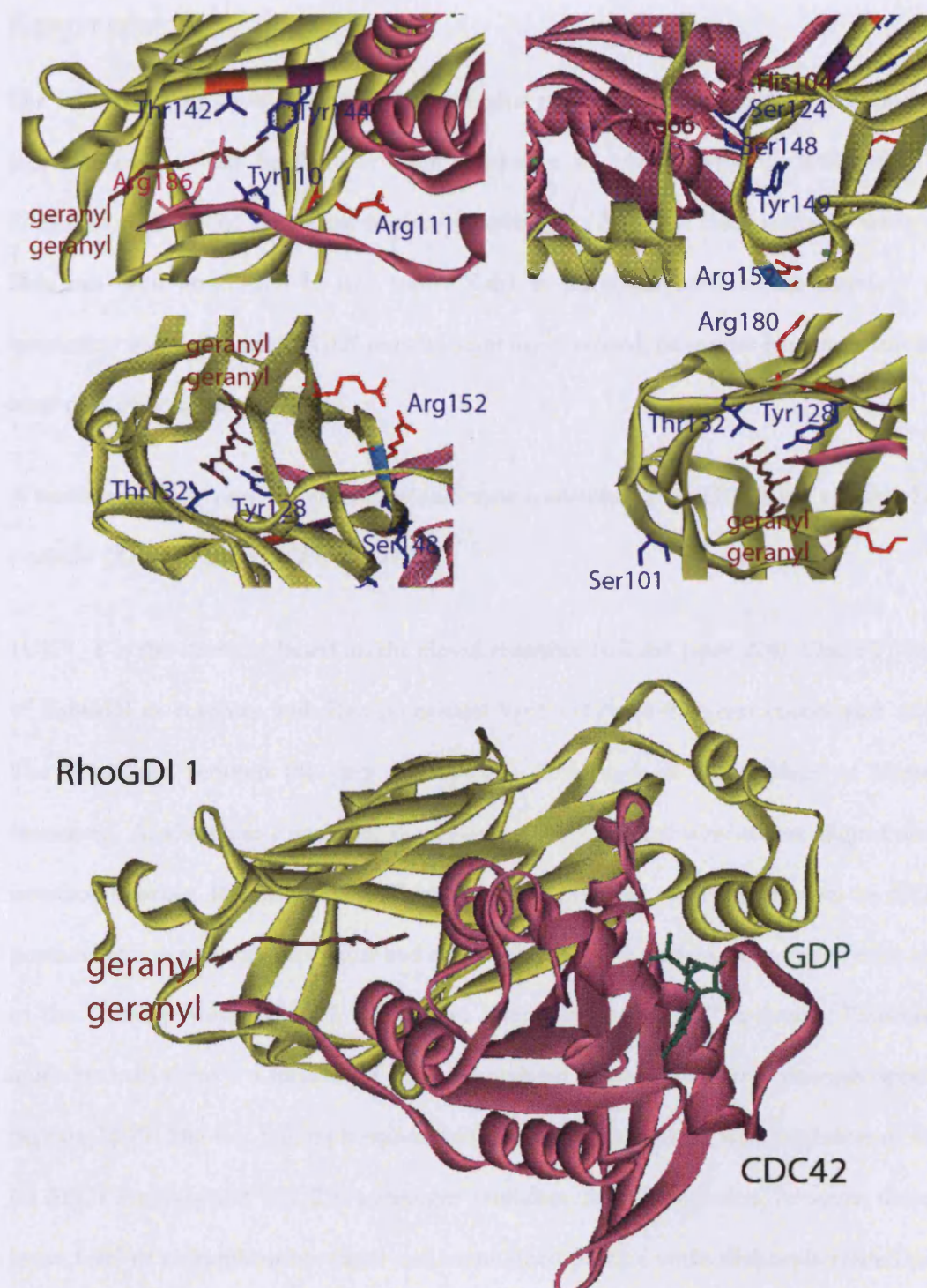
Thr142 interacts directly with Arg186 of Cdc42 (PDB 1doa), and in this case as well a phosphorylation might have a strong influence in lowering the binding; Ser148 is bound to have very much the same effect. Tyr149 does not interact directly with the GTPase, but its phosphorylation very likely strengthens the phosphorylation effect of Ser148. A dimethylated arginine was identified at the C-terminal of the corresponding tryptic peptide.

Interesting modifications were also observed in a different portion of the protein. Tyr110 is tightly inserted in the geranyl-geranyl pocket and its phosphorylation might inhibit the entrance with the geranyl-geranyl group and the binding as a consequence. Very interestingly, next to this phosphorylation site dimethylated Arg111 was identified. This is pointing to the surface but is over 7Å from the nearest RhoGDI-1 atom. Ser101 was already known as a functional phosphorylation site in RhoGDI-1, although the position in the structure does not appear to be immediately related to the interaction between the two proteins.

A set of phosphorylation was found to be switched on after stimulation, and was specific for one of the two isoforms of RhoGDI. Ser124 is very close to the interface with the Rho protein, and establishes a long interaction with Arg66 of Cdc42. Phosphorylation on Ser124 most certainly changes the affinity of the binding, although it is not possible to predict in which direction. Supporting the importance of Ser124 interaction with Arg66, it has been shown that R66E Cdc42 mutants fail to form a complex with GDI (Gibson et al. 2001).

Tyr128 and Thr132 (better candidates than Thr133) synergically bind the geranyl-geranyl of the G-protein bound, and their phosphorylation is likely to push the isoprenic tail towards the surface favouring membrane binding and GDI dissociation.

Modification of this residue might alter the affinity in a non-isoform specific manner. The arginine methylations are located in beta sheets of RhoGDI, are entirely surface exposed, and do not point into the interface. This suggests that methylation of the arginine residues may be controlling the interactions with different phosphokinases and/or other proteins that influence the physiological roles of RhoGDI.



**Figure 57. RhoGDI-1 in complex with Cdc42 structure 1doa**

Geranyl-geranyl is in brown, GDP in green. The top panel represent enlargements of the same structure with the residues of interest annotated. Potentially phosphorylated residues on RhoGDI-1 are in blue, dimethylarginines in red. Interacting residues on Cdc42 are in pink.



### *Rab proteins*

The role of the phosphorylation in Rab proteins regulation is not yet clear. Although many members of the Rab family have been shown to be phosphorylated (Bailly et al. 1991; Fitzgerald et al. 1999), no record of phosphorylation of RabGDI three isoforms are available. This has been confirmed in this study: Rab1 is phosphorylated at the interface of the interaction with GDI, whilst GDI remains unphosphorylated, or at least has a very low level of basal phosphorylation.

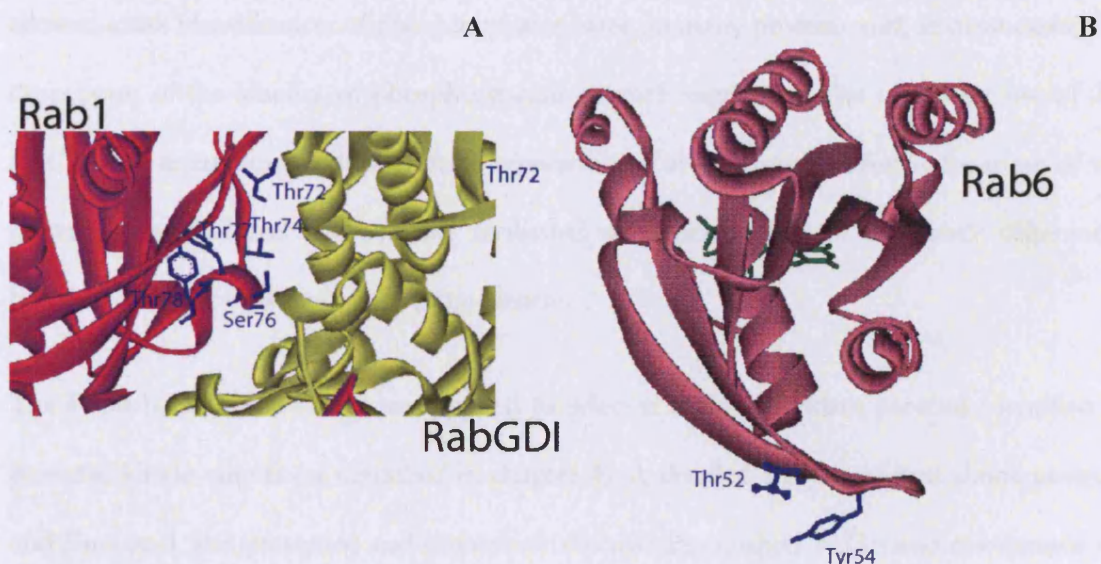
A number of Rab proteins structures and their complexes with GDI have recently become available (An et al. 2003; Rak et al. 2003).

1UKV\_Y is the structure based on the closed sequence to Rab1 (spot 228) (Chain Y, Structure of RabGDI in complex with Rab prenylated Ypt1 GTPases-the yeast counterpart of Rab1). The homology between the two sequences is 81%, and all the residues of interest are conserved. As shown in Figure 58, all possible phosphorylated residues are aligned along the interface between Rab and RabGDI and do not interfere with the binding to GDP. All potential amino acids are accessible and mobile and they are likely to play a role in the strength of the complex with RabGDI rather than interacting with GDP molecule. Particularly, in quiescent cells there is a mixture of diphosphorylated peptide 70-79 and monophosphorylated peptide 72-79. The fact that diphosphorylation might be connected with inhibition of cleavage on Arg71 suggests that Thr72 is a stronger candidate. After stimulation, however, there was a lower level of monophosphorylated and unmodified peptide while diphosphorylated peptides were not present anymore. This suggests that a much higher degree of modification is associated to this part of the protein after stimulation. Indeed, there are two serines, two threonines and two tyrosines packed in a very short peptide (8 residues). Amongst those, only T74 lies in a consensus sequence for PKB (protein sequence was screened using Motifscan

programme. All possible phosphorylation sites have all been highlighted on the sequence (Figure 58, A).

## DISCUSSION

The structure 1D5C\_A (Crystal Structure of Rab6 complexed with GDP) was the most representative for Rab6, the homology between the sequences being 84%. Thr54 and Tyr56 are candidates for the monophosphorylation on Rab6; the kinetics related to this modification were not clear, and implied that more modification on the same peptide might occur. Interestingly, upon stimulation only 10% on the peptide 54-60 and only 70% of the monophosphorylated remain unmodified. This suggests that there might be dephosphorylation on the same peptide, but no corresponding peaks could be found in the MALDI-TOF spectrum. It appears that the candidate residues Thr52 and Tyr54 are both very accessible, but the role of phosphorylation is not obvious because they lie far from the binding site and from the GDP (Figure 58, C).



**Figure 58. Rab1 and Rab6 structures**

(A) Rab1 complex with RabGDI structure 1UKV\_Y. Residues Thr72, Thr74, Tyr77, Tyr78, Ser75, and Ser76 are highlighted as potential phosphorylation sites in regulating the binding upon EGF stimulus. GDI protein is in yellow; (B) Rab6 structure 1D5C\_A. Residues Thr52 and Tyr54 highlighted. GDP is in green, Prenylation is in pink.

## **CHAPTER 7**

### **DISCUSSION**

#### **Summary**

The work presented in this thesis is an original proteomics approach to the study of EGF signalling. First, a reliable method for enriching the phosphoproteome has been established (chapter 3). Second, a subset of phosphoproteins of potential target for some kinases super families was identified. In order to understand the role of these proteins a body of additional information was needed, including the identification of the phosphoproteins of interest and the allocation of their phosphorylation sites.

In order to complement this qualitative data with quantitative information on each phosphorylation site, proteomics technology currently in use was stretched by creating a new combined method, called the Phospho-SILAC strategy. The application of this strategy allowed exact identification of phosphorylation sites on many proteins and, in most cases, the description of the kinetics of phosphorylation on each single site. The consistent use of 2D PAGE as a separation method allowed separation of the phospho-isoforms for most of the proteins identified and consequently facilitated an understanding of the subtle differences between different isoforms of the same protein.

The Phospho-SILAC strategy was applied to selected highly abundant proteins identified as potential kinase targets (as described in chapter 4). A detailed analysis of heat shock proteins and Enolase-1 was presented and extensively discussed in chapter 5. General conclusions on the behaviour of highly abundant phosphoproteins are expressed in this section.

In order to offer a more comprehensive quantitative analysis of the phosphorylation events which occur in the cell upon EGF stimulation, the Phospho-SILAC strategy was also applied to the whole phosphoproteome. The most abundant proteins were selected in order to obtain a more complete set of quantitative data. The analysis of these data led to general conclusions concerning the dynamics of phosphorylation events connected to EGF, with special focus on small GTPases and their effectors. An interesting model of the regulation of some GTPases by post translation modification is offered in this chapter.

## **The limitations of the Phospho-SILAC strategy**

Some difficulties were encountered in the application of the Phospho-SILAC strategy. These have influenced the results obtained with this approach. The specific limitations of each of the techniques used have been discussed in the corresponding chapters (3, 4 and 5) and are summarised here with suggestions of possible ways to overcome them.

A chromatographic procedure has been developed to enrich phosphoproteins based on IMAC. The classic limitations, such as poor protein solubility at low pHs and interference of acidic proteins which were encountered in the earlier attempts, were overcome (chapter 3). However, the leakage of a few proteins containing  $^{32}\text{P}$  (assumed to be entirely phosphoproteins) could not be avoided. At the same time, very few spots which did not contain  $^{32}\text{P}$  were systematically found in each gel. Nevertheless, a sophisticated bioinformatics analysis demonstrated that acidic proteins, typical contaminants for IMAC, were not systematically isolated. This proved that the specificity of the enrichment was high. On the whole, the limitations described did not influence the quality of the enrichment, because most of the phosphoproteins were well separated and could be easily selected from the gel.

The SILAC tagging has been described in chapter 1 and in the introduction to chapter 5. Although SILAC is one of the most recognised techniques for the relative quantification of proteins it contains some inherent boundaries, which become even more relevant when applied to the quantification of single peptides.

Fundamentally, only peptides containing one or more arginines can be quantified, and this constitutes on average ca. 50% of the total tryptic peptides obtained from a protein. This is particularly relevant in the performed experiments because phosphorylated peptides lacking the arginine were sometimes identified. In these cases it was not possible to correlate the phosphorylation event with the EGF stimulus.

Another limitation in the analysis of labelled proteins is that there might be overlapping doublets. Although this is a problem mainly restricted to very abundant proteins because of the corresponding spectra being very rich in signals, it still prevented the quantification of some of the peptides.

Normalisation of the intensities in the MALDI-TOF spectra was necessary for every experiment. Small changes in the amounts of protein lysates loaded on the gel needed to be compensated by recalibration of the ratio of the doublets. This was possible because it was assumed that there are no changes in protein expression within the first two minutes of stimulation and, as a consequence the ratio of corresponding unmodified peptides must be 1:1. However, manual processing was essential because some unmodified peptides (which had been found to be modified as well) had to be excluded from the calibration. These peptides were used instead for indirect quantification (chapter 5, introduction) because their ratio compensated for the ratio of the corresponding modified peptides.

The main disadvantage of the strategy presented is the lack of appropriate software that would recognise doublets and calculate the ratio of their intensities. The entire analysis was carried out by hand, thus increasing the awareness of the requirements that new software must meet. It is essential to determine a priority in the attribution of the peaks from a MALDI-TOF spectrum as the allowance of a high degree of modified residues, such as arginine labelling and serine, threonine and tyrosine phosphorylation leads to a higher level of non-specific attributions. In order to limit this problem, database searches were always performed first allowing only labelled arginine as a possible modification. Ideally, an appropriate software application should be able to recognise doublets not only on the basis of the differences in masses between two peaks but also on the isotopic profile of the two, which is expected to be identical (for identical peptides). Another point to be considered is that peptides containing oxidised methionine can hardly give a real estimation of the relative abundance of the two peptides at different times because the process of oxidation happens during sample processing and is not necessarily the same for the two labelled and unlabelled peptides (in fact, mixtures of oxidised and unmodified methionines containing peptides were often found).

Furthermore, some typical limitations of the majority of proteomics approaches could not be overcome in the presented approach. It is known that when proteins are separated on 2D gel the most abundant proteins will give the strongest spots and will be the easiest to identify. Low abundant proteins often remain obscured. By enriching the phosphoprotein fraction, this problem is only partially overcome. As phosphorylation is a very common modification, the complexity of the sample is only relatively reduced when phosphoproteome is isolated. In fact, a large fraction of highly abundant proteins was identified. However, this is a double edged sword. Highly abundant proteins are usually structural or multi-functional proteins. It is reasonable to think that multi-tasking proteins tend to be modified to a very high degree, probably in order to discriminate amongst the different roles they play in time and space.

Fortunately, the analysis of abundant proteins was often very thorough, especially when more isoforms were involved. For this reasons abundant proteins must not be dismissed, and what has always been viewed as a limitation of the proteomics approach might actually be regarded as a favourable factor. Indeed, abundant proteins proved to be a valuable source of information not otherwise available at the genome level.

## **General observations on phosphorylation dynamics in the cell**

The Phospho-SILAC method was applied to a wider range of proteins in order to investigate if there is a common behaviour in different classes of proteins upon stimulation with EGF. 56 spots have been analysed and 24 different proteins have been identified. As many as 13 of these proteins presented multiple phospho-isoforms (chapter 6). As expected, many proteins do not present any changes upon stimulation with EGF. In the fraction chosen 40% show some kind of change, this percentage including proteins previously known to be changing their phosphorylation state upon EGF stimulation (chapter 4).

Phospho-isoforms are often present in a typical train pattern on the 2D gel where each spot corresponds to the same protein (same primary structure) with a different degree of phosphorylation. Phosphorylation sites were frequently in common with at least two of the isoforms, but their kinetics were different. These isoforms could be described as a population of the same protein that becomes phosphorylated at different rates. For instance, one of Hsp8 isoforms was not phosphorylated on Thr163, one in ten molecules of another isoform became phosphorylated after 5 minutes, and up to three in ten molecules of a third isoform did the same. It looks as though these isoforms become phosphorylated in a wave, maybe to modulate and keep the process under control (chapter 5). To further support this, the Hsp8 isoforms

which became immediately dephosphorylated before two minutes of EGF stimulation are the most acidic isoforms, and therefore likely the most phosphorylated; the proteins at higher pI followed the same process though more slowly (chapter 4). Subsequent results suggested that this dephosphorylation occurs on Thr38 (chapter 5).

Interestingly, similar observations can be made for Enolase-1: less than 10% for one Enolase-1 isoform became phosphorylated whilst up to 40% of another phospho-isoform became phosphorylated on the same residue.

This behaviour is probably biologically significant rather than caused by the experimental processing of the cells; in fact, the cells are synchronised in order to exclusively connect the response to EGF stimulation. Although the cells are resting and quiescent, a distribution of different phospho-isoforms of the same proteins is already present. This automatically leads to a differentiated response of these isoforms to the stimulus. These phospho-isoforms are most likely the result of protein functional partitioning. These events could be seen as a regulatory process of the cell which does not recruit all its proteins in the response to one stimulus, but keeps some inactive (or only partially activated) for other possible functions. Another process that became particularly evident, and that is in line with this latter deduction, is that the stoichiometry of the phosphorylation is very rarely 100%, but rather between 10-50%.

## **Chaperones involved in EGF signalling**

In chapter 4 an immuno screening for serine/threonine phosphorylation of the total phosphoproteome was described. The phosphorylated serines and threonines selected for this analysis are in a well known consensus motif recognised by proteins kinases. However, due to the inherent limitations of this approach, mainly chaperones and structural proteins were found. In the discussion to chapter 4, evidence was given to show that each of the highly



abundant proteins found in the screening is strongly related to stress induced signalling. These observations led to the conclusion that phosphorylation and oxidation are strongly related in signalling. It is not easy to hypothesise on the molecular mechanisms mediating this connection, especially because they are not necessarily direct, but might be finely regulated through a net of different isoforms. The oxidation and reduction of protein cysteine sulfhydryl (SH) groups work as a molecular switch to start or stop signalling in combination with phosphorylation. It is known that oxidation of cysteine SH groups on protein tyrosine phosphatases switches off the action of protein tyrosine phosphatases (Cho et al. 2004). This event may not, however, signal for initial autophosphorylation of previously unphosphorylated PTKs, whereas it certainly prevents dephosphorylation of once-phosphorylated PTKs. New mechanisms for oxidative stress-mediated PTK activation assume that intracellular specific cysteine SH groups on PTK proteins can be another target of oxidative stress for inducing a conformational change necessary for initial activation of PTKs (Nakashima et al. 2002). Considering that redox events trigger phosphorylation events, it can be suggested that phosphorylation/dephosphorylation of the chaperones and other proteins involved in the redox network influences the stability of the chaperone substrate-binding form, therefore regulating the redox reactions. As a consequence, an initial phosphorylation of some chaperones upon growth factor stimulus passes on the activation of kinases and phosphatases to propagate the signal further. Although this might not be the key path for the response to EGF signal, it could well support the other better known cascades.

These results, together with previous evidence (see chapter 1) hinted that chaperones play a significant role, possibly complementary to the one of other signalling proteins. For this reason, Hsp70/Hsc70 family was successfully analysed to further understand which phosphorylations are involved in their regulation and to obtain quantitative data on potential phosphorylation on TP, as observed in the immuno screening. Indeed, it was concluded that

Hsp8 is phosphorylated on Thr 38 (followed by proline). A model was proposed according to which this phosphorylation regulates the hydrolysis of ATP and the chaperone activity itself (refer to discussion of chapter 5). The investigation was extended to more chaperones belonging to the same family in order to identify some common behaviour, reaching the conclusion that different proteins belonging to the same superfamily of Hsc70 are phosphorylated mostly on conserved residues. A specific case was highlighted, Hsp8 and Hsp5, which share only 63% homology, have one phosphorylated residue in common (Thr 163-on Hsp8 sequence), although this follows different phosphorylation kinetics (for details, see Discussion to chapter 6).

Hsp8 forms 1:1 complexes with Stress induced phosphoprotein 1 which are regulated by phosphorylation events (van der Spuy et al. 2001). It is difficult to understand which isoforms of these proteins are involved. Although phosphorylation on Stress induced phosphoprotein 1 was further analysed using the Phospho-SILAC method, it showed no modifications.

## **Rab/RabGDI vs. Rho/RhoGDI-1 interactions**

Combining the results obtained from the analysis of the two Rab proteins and RhoGDI-1 together with data available from literature, some new observations can be made. Rho/RhoGDI-1 protein interaction appears mostly regulated by phosphorylation on GDI, whilst Rab/RabGDI is regulated via phosphorylation on Rabs. These two complexes regulate different biological processes, but are mechanistically very similar. The data show that RhoGDI-1 presented some phosphorylations on residues interacting with the geranyl-geranyl or the GTPase bound to it. At the same time it also has phosphorylation sites along the surface that interact with the GTPase's amino acidic chain. It seems that the two sets of phosphorylation move in the same direction, which is dissociation for the GTPase and exposure of the geranyl-geranyl group. Dissociation of the GDI molecules causes exposure of

the geranyl-geranyl group, resulting in the GTPase's binding to the membrane and consequently triggering the signal response. Some data suggest that phosphorylation on the G-protein increases the affinity of the RhoA and Cdc42 with RhoGDI. This suggests that at the same time phosphorylation on RhoGDI-1 lowers such affinity.

As for the Rab/RabGDI beta complex, opposite conclusions to the ones reached for Rho/RhoGDI-1 complex could be made. Both Rab6 and Rab1 were identified and phosphorylation changes were found for both. For Rab1 in particular, it appears obvious that such phosphorylations interfere with the affinity of the complex. On the other hand, RabGDI beta was identified and half of its sequence covered, but it was not possible to find any putative sites. They are likely to be present, but are probably not as abundant as in RhoGDI.

## **Conclusions and outlook**

EGF signalling was investigated using an innovative proteomics approach. Looking at the larger frame of proteomics (Stannard et al. 2004) it becomes clear that revealing signalling pathways and cellular functions at a molecular level is far more complex than previously anticipated. This is due to function partitioning of the proteins, their multitasking and moonlighting and there is evidence that this accounts for isoforms and splicing variants (Godovac-Zimmermann et al. 2005; Kleiner et al. 2005; Stannard et al. 2004). The analysis presented in this thesis provides numerous leads to understand the role of the phosphorylations observed, some supported by biological experiments and structural observations. More sophisticated information has been obtained here concerning EGF signalling than has been achieved in previous proteomics experiments. Above all, the importance of quantitative proteomics in the study of EGF signalling has been demonstrated and the information obtained can be used and built upon in further investigations.

It is often believed that advanced current proteomics methods are readily available. This is usually true for micro-organisms, but certainly not for eukaryotes. This needs to be addressed by the further development of methods for MS/MS of proteins rather than peptides as protein surrogates, as this analysis lacks information on isoforms. These improvements need to involve methods for protein isoform isolations and protein MS in order to achieve new information on protein-protein interaction and complexes in cellular signalling.

The logical steps based on the work explained in this thesis would include a set of molecular biology experiments in which the amino acids described here would be mutated in order to understand their role in the cell, or simply in order to confirm the predictions made here based on general observations on phosphorylation sites and their locations on the structures. This would certainly contribute further to the knowledge of EGF signalling.

## REFERENCES

Adam, G. C., Sorensen, E. J., & Cravatt, B. F. 2002, "Chemical strategies for functional proteomics", *Mol.Cell Proteomics*, vol. 1, no. 10, pp. 781-90.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor Miklos, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., Pablos, B. d., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., & Venter, J. C. 2000, "The Genome Sequence of *Drosophila melanogaster*", *Science*, vol. 287, no. 5461, pp. 2185-2195.

Adler, V., Yin, Z., Fuchs, S. Y., Benezra, M., Rosario, L., Tew, K. D., Pincus, M. R., Sardana, M., Henderson, C. J., Wolf, C. R., Davis, R. J., & Ronai, Z. 1999, "Regulation of JNK signaling by GSTp", *EMBO J.*, vol. 18, no. 5, pp. 1321-1334.

Aebersold, R. & Goodlett, D. R. 2001, "Mass spectrometry in proteomics", *Chem.Rev.*, vol. 101, no. 2, pp. 269-295.

Aebersold, R. & Mann, M. 2003, "Mass spectrometry-based proteomics", *Nature*, vol. 422, no. 6928, pp. 198-207.

- An, Y., Shao, Y., Alory, C., Matteson, J., Sakisaka, T., Chen, W., Gibbs, R. A., Wilson, I. A., & Balch, W. E. 2003, "Geranylgeranyl Switching Regulates", *Structure*, vol. 11, no. 3, pp. 347-357.
- Andersson, L. & Porath, J. 1986, "Isolation of phosphoproteins by immobilized metal (Fe<sup>3+</sup>) affinity chromatography", *Anal.Biochem.*, vol. 154, no. 1, pp. 250-254.
- Annan, R. S. & Carr, S. A. 1996, "Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry", *Anal.Chem.*, vol. 68, no. 19, pp. 3413-3421.
- Arakawa, T. & Timasheff, S. N. 1985, "Theory of protein solubility", *Meth.Enzymol.*, vol. 114, pp. 49-77.
- Arimura, N., Menager, C., Fukata, Y., & Kaibuchi, K. 2004, "Role of CRMP-2 in neuronal polarity", *J.Neurobiol.*, vol. 58, no. 1, pp. 34-47.
- Bailly, E., McCaffrey, M., Touchot, N., Zahraoui, A., Goud, B., & Bornens, M. 1991, "Phosphorylation of two small GTP-binding proteins of the Rab family by p34cdc2", *Nature*, vol. 350, no. 6320, pp. 715-718.
- Barbieri, M. A., Fernandez-Pol, S., Hunker, C., Horazdovsky, B. H., & Stahl, P. D. 2004, "Role of rab5 in EGF receptor-mediated signal transduction", *Eur.J.Cell Biol*, vol. 83, no. 6, pp. 305-314.
- Baselga, J. 2002, "Why the Epidermal Growth Factor Receptor? The Rationale for Cancer Therapy", *Oncologist*, vol. 7, no. 90004, pp. 2-8.
- Baty, J. W., Hampton, M. B., & Winterbourn, C. C. 2005, "Proteomic detection of hydrogen peroxide-sensitive thiol proteins in Jurkat cells", *Biochem.J.*, vol. 389, no. Pt 3, pp. 785-795.
- Berk, B. C., Brock, T. A., Webb, R. C., Taubman, M. B., Atkinson, W. J., Gimbrone, M. A., Jr., & Alexander, R. W. 1985, "Epidermal growth factor, a vascular smooth muscle mitogen, induces rat aortic contraction", *J.Clin.Invest.*, vol. 75, no. 3, pp. 1083-1086.
- Bimston, D., Song, J., Winchester, D., Takayama, S., Reed, J., & Morimoto, R. 1998, "BAG-1, a negative regulator of Hsp70 chaperone activity, uncouples nucleotide hydrolysis from substrate release", *EMBO J.*, vol. 17, no. 23, pp. 6871-6878.
- Blagoev, B., Ong, S. E., Kratchmarova, I., & Mann, M. 2004, "Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics", *Nat.Biotechnol.*, vol. 22, no. 9, pp. 1139-45.
- Boado, R. J., Campbell, D. A., & Chopra, I. J. 1988, "Nucleotide sequence of rat liver iodothyronine 5'-monodeiodinase (5' MD): its identity with the protein disulfide isomerase", *Biochem.Biophys.Res.Comm.*, vol. 155, no. 3, pp. 1297-1304.

Bogdanov, B. & Smith, R. D. 2004, "Proteomics by FTICR mass spectrometry: Top down and bottom up", *Mass Spectrom.Rev.*, vol. 30, p. 30.

Boni-Schnetzler, M. & Pilch, P. F. 1987, "Mechanism of epidermal growth factor receptor autophosphorylation and high-affinity binding", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 84, no. 22, pp. 7832-7836.

Bourmeyster, N. & Vignais, P. V. 1996, "Phosphorylation of Rho GDI Stabilizes the Rho A-Rho GDI Complex in Neutrophil Cytosol", *Biochem.Biophys.Res.Comm.*, vol. 218, no. 1, pp. 54-60.

Briknarova, K., Takayama, S., Brive, L., Havert, M. L., Knee, D. A., Velasco, J., Homma, S., Cabezas, E., Stuart, J., Hoyt, D. W., Satterthwait, A. C., Llinas, M., Reed, J. C., & Ely, K. R. 2001, "Structural analysis of BAG1 cochaperone and its interactions with Hsc70 heat shock protein", *Nat.Struct.Biol.*, vol. 8, no. 4, pp. 349-352.

Brive, L. & Abagyan, R. 2002, "Computational structural proteomics", *Ernst.Schering.Res.Found.Workshop* no. 38, pp. 149-166.

Bukau, B. & Horwich, A. L. 1998, "The Hsp70 and Hsp60 chaperone machines", *Cell*, vol. 92, no. 3, pp. 351-366.

Burgering, B. M., Vries-Smits, A. M., Medema, R. H., van Weeren, P. C., Tertoolen, L. G., & Bos, J. L. 1993, "Epidermal growth factor induces phosphorylation of extracellular signal-regulated kinase 2 via multiple pathways", *Mol.Cell Biol.*, vol. 13, no. 12, pp. 7248-7256.

Cantley, L. C. 2002, "The phosphoinositide 3-kinase pathway", *Science*, vol. 296, no. 5573, pp. 1655-1657.

Chaga, G. S. 2001, "Twenty-five years of immobilized metal ion affinity chromatography: past, present and future", *Biochem.Biophys.Methods*, vol. 49, no. 1-3, pp. 313-334.

Chai, G., Brewer, J. M., Lovelace, L. L., Aoki, T., Minor, W., & Lebioda, L. 2004, "Expression, Purification and the 1.8 Å Resolution Crystal Structure of Human Neuron Specific Enolase", *J.Mol.Biol.*, vol. 341, no. 4, pp. 1015-1021.

Chen, W., Daines, M. O., & Hershey, G. K. K. 2004, "Methylation of STAT6 Modulates STAT6 Phosphorylation, Nuclear Translocation, and DNA-Binding Activity", *J.Immunol.*, vol. 172, no. 11, pp. 6744-6750.

Chevalier, S., Macdonald, N., Tonge, R., Rayner, S., Rowlinson, R., Shaw, J., Young, J., Davison, M., & Roberts, R. A. 2000, "Proteomic analysis of differential protein expression in primary hepatocytes induced by EGF, tumour necrosis factor alpha or the peroxisome proliferator nafenopin", *Eur.J.Biochem.*, vol. 267, no. 15, pp. 4624-4634.

- Cho, S. H., Lee, C. H., Ahn, Y., Kim, H., Kim, H., Ahn, C. Y., Yang, K. S., & Lee, S. R. 2004, "Redox regulation of PTEN and protein tyrosine phosphatases in H<sub>2</sub>O<sub>2</sub> mediated cell signaling", *FEBS Lett.*, vol. 560, no. 1-3, pp. 7-13.
- Cohen, S. 1962, "Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal", *J.Biol.Chem.*, vol. 237, pp. 1555-1562.
- Coin, F., Auriol, J., Tapias, A., Clivio, P., Vermeulen, W., & Egly, J. M. 2004, "Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity", *EMBO J.*, vol. 23, no. 24, pp. 4835-4846.
- Connolly, J. L., Green, S. A., & Greene, L. A. 1984, "Comparison of rapid changes in surface morphology and coated pit formation of PC12 cells in response to nerve growth factor, epidermal growth factor, and dibutyryl cyclic AMP", *J.Cell Biol.*, vol. 98, no. 2, pp. 457-465.
- Cottrell, J. S. 1994, "Protein identification by peptide mass fingerprinting", *Pept.Res.*, vol. 7, no. 3, pp. 115-24.
- Cowan, K. J. & Storey, K. B. 2003, "Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress", *J.Exp.Biol.*, vol. 206, no. 7, pp. 1107-1115.
- Cunnick, J. M., Meng, S., Ren, Y., Despons, C., Wang, H. G., Djeu, J. Y., & Wu, J. 2002, "Regulation of the mitogen-activated protein kinase signaling pathway by SHP2", *J.Biol.Chem.*, vol. 277, no. 11, pp. 9498-9504.
- Daub, H., Wallasch, C., Lankenau, A., Herrlich, A., & Ullrich, A. 1997, "Signal characteristics of G-protein-transactivated EGF receptor", *EMBO J.*, vol. 16, no. 23, pp. 7032-7044.
- David, M., Wong, L., Flavell, R., Thompson, S. A., Wells, A., Larner, A. C., & Johnson, G. R. 1996, "STAT activation by epidermal growth factor (EGF) and amphiregulin. Requirement for the EGF receptor kinase but not for tyrosine phosphorylation sites or JAK1", *J.Biol.Chem.*, vol. 271, no. 16, pp. 9185-9188.
- De Leenheer, A. P. & Thienport, L. M. 1992, "Application of isotope dilution-mass spectrometry in clinical chemistry, pharmacokinetics, and toxicology", *Mass Spectrom.Rev.*, vol. 11, pp. 249-307.
- DerMardirossian, C., Schnelzer, A., & Bokoch, G. M. 2004, "Phosphorylation of RhoGDI by Pak1 mediates dissociation of Rac GTPase", *Mol.Cell.*, vol. 15, no. 1, pp. 117-127.
- Dittmar, K. D., Banach, M., Galigniana, M. D., & Pratt, W. B. 1998, "The Role of DnaJ-like Proteins in Glucocorticoid Receptor+hsp90 Heterocomplex Assembly by the Reconstituted hsp90+p60+hsp70 Foldosome Complex", *J.Biol.Chem.*, vol. 273, no. 13, pp. 7358-7366.



Domagala, T., Konstantopoulos, N., Smyth, F., Jorissen, R. N., Fabri, L., Geleick, D., Lax, I., Schlessinger, J., Sawyer, W., & Howlett et, a. 2000, "Stoichiometry, kinetic and binding analysis of the interaction between epidermal growth factor (EGF) and the extracellular domain of the EGF receptor", *Growth Factors*, vol. 18, no. 1, pp. 11-29.

Doong, H., Price, J., Kim, Y. S., Gasbarre, C., Probst, J., Liotta, L. A., Blanchette, J., Rizzo, K., & Kohn, E. 2000, "CAIR-1/BAG-3 forms an EGF-regulated ternary complex with phospholipase C-gamma and Hsp70/Hsc70", *Oncogene*, vol. 19, no. 38, pp. 4385-4395.

Droge, W. 2002, "Aging-related changes in the thiol/disulfide redox state: implications for the use of thiol antioxidants", *Exp.Gerontol.*, vol. 37, no. 12, pp. 1333-1345.

Eguchi, S., Numaguchi, K., Iwasaki, H., Matsumoto, T., Yamakawa, T., Utsunomiya, H., Motley, E. D., Kawakatsu, H., Owada, K. M., Hirata, Y., Marumo, F., & Inagami, T. 1998, "Calcium-dependent Epidermal Growth Factor Receptor Transactivation Mediates the Angiotensin II-induced Mitogen-activated Protein Kinase Activation in Vascular Smooth Muscle Cells", *J.Biol.Chem.*, vol. 273, no. 15, pp. 8890-8896.

Eng, J. K., McCormack, A. L., & Yates, I. J. R. 1994, "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database", *J.Am.Soc.Mass Spectrom.*, vol. 5, no. 11, pp. 976-989.

Engel, M., Mazurek, S., Eigenbrodt, E., & Welter, C. 2004, "Phosphoglycerate Mutase-derived Polypeptide Inhibits Glycolytic Flux and Induces Cell Growth Arrest in Tumor Cell Lines", *J.Biol.Chem.*, vol. 279, no. 34, pp. 35803-35812.

Etienne-Manneville, S. & Hall, A. 2003, "Cdc42 regulates GSK-3[beta] and adenomatous polyposis coli to control cell polarity", *Nature*, vol. 421, no. 6924, pp. 753-756.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. 1989, "Electrospray ionization for mass spectrometry of large biomolecules", *Science*, vol. 246, no. 4926, pp. 64-71.

Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H. S., Leahy, D. J., & Lemmon, M. A. 2003, "EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization", *Mol.Cell*, vol. 11, no. 2, pp. 507-517.

Ficarro, S. B., McClelland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., & White, F. M. 2002, "Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*", *Nat.Biotechnol.*, vol. 20, no. 3, pp. 301-5.

Fitzgerald, M. L. & Reed, G. L. 1999, "Rab6 is phosphorylated in thrombin-activated platelets by a protein kinase C-dependent mechanism: effects on GTP/GDP binding and cellular distribution", *Biochem.J.*, vol. 342 ( Pt 2), pp. 353-360.

Fratelli, M., Demol, H., Puype, M., Casagrande, S., Eberini, I., Salmona, M., Bonetto, V., Mengozzi, M., Duffieux, F., Miclet, E., Bachì, A., Vandekerckhove, J., Gianazza, E., & Ghezzi,

P. 2002, "Identification by redox proteomics of glutathionylated proteins in oxidatively stressed human T lymphocytes", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 99, no. 6, pp. 3505-3510.

Fukunaga, R. & Hunter, T. 1997, "MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates", *EMBO J.*, vol. 16, no. 8, pp. 1921-1933.

Gaballo, A., Zanotti, F., & Papa, S. 2002, "Structures and interactions of proteins involved in the coupling function of the protonmotive F(o)F(1)-ATP synthase", *Curr.Protein Pept.Sci.*, vol. 3, no. 4, pp. 451-460.

Gaberc-Porekar, V. & Menart, V. 2001, "Perspectives of immobilized-metal affinity chromatography", *Journal of Biochemical and Biophysical Methods*, vol. 49, no. 1-3, pp. 335-360.

Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H. J., Walker, F., Frenkel, M. J., & Hoyne et, a. 2002, "Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha", *Cell*, vol. 110, no. 6, pp. 763-773.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., & Superti-Furga, G. 2002, "Functional organization of the yeast proteome by systematic analysis of protein complexes", *Nature*, vol. 415, no. 6868, pp. 141-147.

Ge, Y., Rajkumar, L., Guzman, R. C., Nandi, S., Patton, W. F., & Agnew, B. J. 2004, "Multiplexed fluorescence detection of phosphorylation, glycosylation, and total protein in the proteomic analysis of breast cancer refractoriness", *Proteomics*, vol. 4, no. 11, pp. 3464-3467.

Gear, A. R., Simon, C. G., & Polanowska-Grabowska, R. 1997, "Platelet adhesion to collagen activates a phosphoprotein complex of heat-shock proteins and protein phosphatase 1", *J.Neural Transm.*, vol. 104, no. 10, pp. 1037-1047.

Ghosh, A. K., Steele, R., & Ray, R. B. 1999, "MBP-1 Physically Associates with Histone Deacetylase for Transcriptional Repression", *Biochem.Biophys.Res.Comm.*, vol. 260, no. 2, pp. 405-409.

Gibson, R. M. & Wilson-Delfosse, A. L. 2001, "RhoGDI-binding-defective mutant of Cdc42Hs targets to membranes and activates filopodia formation but does not cycle with the cytosol of mammalian cells", *Biochem.J.*, vol. 359, no. Pt 2, pp. 285-294.

Godovac-Zimmermann, J. & Brown, L. R. 2003, "Proteomics approaches to elucidation of signal transduction pathways", *Curr.Opin.Mol.Ther.*, vol. 5, no. 3, pp. 241-249.

- Godovac-Zimmermann, J., Kleiner, O., Brown, L. R., & Drukier, A. K. 2005, "Perspectives in spicing up proteomics with splicing", *Proteomics*, vol. 3, p. 3.
- Gorvel, J. P., Chang, T. C., Boretto, J., Azuma, T., & Chavrier, P. 1998, "Differential properties of D4/LyGDI versus RhoGDI: phosphorylation and rho GTPase selectivity", *FEBS Lett.*, vol. 422, no. 2, pp. 269-273.
- Goshima, Y., Nakamura, F., Strittmatter, P., & Strittmatter, S. M. 1995, "Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33", *Nature*, vol. 376, no. 6540, pp. 509-514.
- Gosser, Y. Q., Nomanbhoy, T. K., Aghazadeh, B., Manor, D., Combs, C., Cerione, R. A., & Rosen, M. K. 1997, "C-terminal binding domain of Rho GDP-dissociation inhibitor directs N-terminal inhibitory peptide to GTPases", *Nature*, vol. 387, no. 6635, pp. 814-819.
- Graus-Porta, D., Beerli, R. R., Daly, J. M., & Hynes, N. E. 1997, "ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling", *EMBO J.*, vol. 16, no. 7, pp. 1647-1655.
- Gregory, H. 1975, "Isolation and structure of urogastrone and its relationship to epidermal growth factor", *Nature*, vol. 257, no. 5524, pp. 325-327.
- Grizot, S., Faure, J., Fieschi, F., Vignais, P. V., Dagher, M. C., & Pebay-Peyroula, E. 2001, "Crystal structure of the Rac1-RhoGDI complex involved in nadph oxidase activation", *Biochemistry*, vol. 40, no. 34, pp. 10007-10013.
- Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., Ohara, O., Mann, M., Jensen, O. N., & Pandey, A. 2002, "A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate", *Mol. Cell Proteomics*, vol. 1, no. 7, pp. 517-27.
- Gu, Y., Hamajima, N., & Ihara, Y. 2000, "Neurofibrillary tangle-associated collapsin response mediator protein-2 (CRMP-2) is highly phosphorylated on Thr-509, Ser-518, and Ser-522", *Biochemistry*, vol. 39, no. 15, pp. 4267-4275.
- Guerrera, I. C. & Kleiner, O. 2005a, "Application of Mass Spectrometry in Proteomics", *Biosci. Rep.*, vol. in press.
- Guerrera, I. C., Predic-Atkinson, J., Kleiner, O., Soskic, V., & Godovac-Zimmermann J. 2005b, "Enrichment of Phosphoproteins for Proteomic Analysis Using Immobilised Fe(III)-Affinity Adsorption Chromatography", *J. Proteome Res.*, vol. in press.
- Hallberg, B., Rayter, S. I., & Downward, J. 1994, "Interaction of Ras and Raf in intact mammalian cells upon extracellular stimulation", *J. Biol. Chem.*, vol. 269, no. 6, pp. 3913-3916.

Hart, S. R., Waterfield, M. D., Burlingame, A. L., & Cramer, R. 2002, "Factors governing the solubilization of phosphopeptides retained on ferric NTA IMAC beads and their analysis by MALDI TOFMS", *J Am.Soc.Mass Spectrom.*, vol. 13, no. 9, pp. 1042-1051.

Hendershot, L. M., Ting, J., & Lee, A. S. 1988, "Identity of the immunoglobulin heavy-chain-binding protein with the 78,000-dalton glucose-regulated protein and the role of posttranslational modifications in its binding function", *Mol.Cell Proteomics*, vol. 8, no. 10, pp. 4250-4256.

Hernandez-Sotomayor, S. M. & Carpenter, G. 1993, "Non-catalytic activation of phospholipase C-gamma 1 in vitro by epidermal growth factor receptor", *Biochem.J.*, vol. 293 ( Pt 2), pp. 507-511.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., & Tyers, M. 2002, "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry", *Nature*, vol. 415, no. 6868, pp. 180-183.

Hofer, F., Berdeaux, R., & Martin, G. S. 1998, "Ras-independent activation of Ral by a Ca(2+)-dependent pathway", *Curr.Biol.*, vol. 8, no. 14, pp. 839-842.

Hoffman, G. R., Nassar, N., & Cerione, R. A. 2000, "Structure of the Rho family GTP-binding protein Cdc42 in complex with the multifunctional regulator RhoGDI", *Cell*, vol. 100, no. 3, pp. 345-356.

Hohfeld, J., Minami, Y., & Hartl, F. U. 1995, "Hip, a novel cochaperone involved in the eukaryotic Hsc70/Hsp40 reaction cycle", *Cell*, vol. 83, no. 4, pp. 589-598.

Hubbard, S. R., Mohammadi, M., & Schlessinger, J. 1998, "Autoregulatory mechanisms in protein-tyrosine kinases", *J.Biol.Chem.*, vol. 273, no. 20, pp. 11987-11990.

Hunter, T. 1998, "The role of tyrosine phosphorylation in cell growth and disease", *Harvey Lect.*, vol. 94, pp. 81-119.

Hunter, T. 1995, "Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling", *Cell*, vol. 80, no. 2, pp. 225-236.

Hur, E. M. & Kim, K. T. 2002, "G-protein-coupled receptor signalling and cross-talk: achieving rapidity and specificity", *Cell Signal*, vol. 14, no. 5, pp. 397-405.

Ibarrola, N., Molina, H., Iwahori, A., & Pandey, A. 2004, "A novel proteomic approach for specific identification of tyrosine kinase substrates using [<sup>13</sup>C]tyrosine", *J.Biol.Chem.*, vol. 279, no. 16, pp. 15805-13.

Immler, D., Gremm, D., Kirsch, D., Spengler, B., Presek, P., & Meyer, H. E. 1998, "Identification of phosphorylated proteins from thrombin-activated human platelets isolated by two-dimensional gel electrophoresis by electrospray ionization-tandem mass spectrometry (ESI-MS/MS) and liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS)", *Electrophoresis*, vol. 19, no. 6, pp. 1015-1023.

Isomura, M., Kikuchi, A., Ohga, N., & Takai, Y. 1991, "Regulation of binding of rhoB p20 to membranes by its specific regulatory protein, GDP dissociation inhibitor", *Oncogene*, vol. 6, no. 1, pp. 119-124.

Ito, T., Kagoshima, M., Sasaki, Y., Li, C., Uda, N., Kitsukawa, T., Fujisawa, H., Taniguchi, M., Yagi, T., Kitamura, H., & Goshima, Y. 2000, "Repulsive axon guidance molecule Sema3A inhibits branching morphogenesis of fetal mouse lung", *Mech.Dev.*, vol. 97, no. 1-2, pp. 35-45.

Jacob, C., Lancaster, J. R., & Giles, G. I. 2004, "Reactive sulphur species in oxidative signal transduction", *Biochem.Soc.Trans.*, vol. 32, no. Pt 6, pp. 1015-1017.

Jin, W. H., Dai, J., Zhou, H., Xia, Q. C., Zou, H. F., & Zeng, R. 2004, "Phosphoproteome analysis of mouse liver using immobilized metal affinity purification and linear ion trap mass spectrometry", *Rapid Commun.Mass Spectrom.*, vol. 18, no. 18, pp. 2169-2176.

Johnson, G. L. & Vaillancourt, R. R. 1994, "Sequential protein kinase reactions controlling cell growth and differentiation", *Curr.Opin.Cell Biol.*, vol. 6, no. 2, pp. 230-238.

Jones, P. F., Jakubowicz, T., Pitossi, F. J., Maurer, F., & Hemmings, B. A. 1991, "Molecular cloning and identification of a serine/threonine protein kinase of the second-messenger subfamily", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 88, no. 10, pp. 4171-4175.

Kagoshima, M. & Ito, T. 2001, "Diverse gene expression and function of semaphorins in developing lung: positive and negative regulatory roles of semaphorins in lung branching morphogenesis", *Genes Cells*, vol. 6, no. 6, pp. 559-571.

Kang, S., Liao, P. c., Gage, D. A., & Esselman, W. J. 1997, "Identification of in Vivo Phosphorylation Sites of CD45 Protein-tyrosine Phosphatase in 70Z/3.12 Cells", *J.Biol.Chem.*, vol. 272, no. 17, pp. 11588-11596.

Karas, M. & Hillenkamp, F. 1988, "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons", *Anal.Chem.*, vol. 60, no. 20, pp. 2299-301.

Keep, N. H., Barnes, M., Barsukov, I., Badii, R., Lian, L. Y., Segal, A. W., Moody, P. C., & Roberts, G. C. 1997, "A modulator of rho family G-proteins, rhoGDI, binds these G-proteins

via an immunoglobulin-like domain and a flexible N-terminal arm", *Structure*, vol. 5, no. 5, pp. 623-633.

Kim, K. H., Stellmach, V., Javors, J., & Fuchs, E. 1987, "Regulation of human mesothelial cell differentiation: opposing roles of retinoids and epidermal growth factor in the expression of intermediate filament proteins", *J.Cell Biol.*, vol. 105, no. 6 Pt 2, pp. 3039-3051.

Kim, M. K. & Carroll, W. L. 2004, "Autoregulation of the N-myc gene is operative in neuroblastoma and involves histone deacetylase 2", *Cancer*, vol. 101, no. 9, pp. 2106-2115.

Kjeldsen, F., Haselmann, K. F., Budnik, B. A., Sorensen, E. S., & Zubarev, R. A. 2003, "Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage", *Anal.Chem.*, vol. 75, no. 10, pp. 2355-61.

Kleiner, O. & Godovac-Zimmermann, J. 2005, "Proteomics of signal transduction pathways," in *Signalling pathways in liver diseases*, Springer Verlag, pp. 417-431.

Kloth, M. T., Laughlin, K. K., Biscardi, J. S., Boerner, J. L., Parsons, S. J., & Silva, C. M. 2003, "STAT5b, a Mediator of Synergism between c-Src and the Epidermal Growth Factor Receptor", *J.Biol.Chem.*, vol. 278, no. 3, pp. 1671-1679.

Koch, G., Tanaka, K., Masuda, T., Yamochi, W., Nonaka, H., & Takai, Y. 1997, "Association of the Rho family small GTP-binding proteins with Rho GDP dissociation inhibitor (Rho GDI) in *Saccharomyces cerevisiae*", *Oncogene*, vol. 15, no. 4, pp. 417-422.

Kohn, E. A., Yoo, C. J., & Eastman, A. 2003, "The Protein Kinase C Inhibitor Go6976 Is a Potent Inhibitor of DNA Damage-induced S and G2 Cell Cycle Checkpoints", *Cancer Res.*, vol. 63, no. 1, pp. 31-35.

Krebs, E. G. 1994, "The growth of research on protein phosphorylation", *Trends Biochem.Sci.*, vol. 19, no. 11, p. 439.

Kuan, C. T., Wikstrand, C. J., & Bigner, D. D. 2001, "EGF mutant receptor vIII as a molecular target in cancer therapy", *Endocr.Relat.Cancer*, vol. 8, no. 2, pp. 83-96.

Lahaye, D. H., Camps, M. G., Erp, P. E., Peters, P. H., & Zoelen, E. J. 1998, "Epidermal growth factor (EGF) receptor density controls mitogenic activation of normal rat kidney (NRK) cells by EGF", *J.Cell Physiol.*, vol. 174, no. 1, pp. 9-17.

Lakshmikuttyamma, A., Selvakumar, P., Anderson, D. H., Datla, R. S., & Sharma, R. K. 2004, "Molecular cloning of bovine cardiac muscle heat-shock protein 70 kDa and its phosphorylation by cAMP-dependent protein kinase in vitro", *Biochemistry*, vol. 43, no. 42, pp. 13340-13347.

- Langlois, W. J., Sasaoka, T., Saltiel, A. R., & Olefsky, J. M. 1995, "Negative Feedback Regulation and Desensitization of Insulin- and Epidermal Growth Factor-stimulated p21[IMAGE] Activation", *J.Biol.Chem.*, vol. 270, no. 43, pp. 25320-25323.
- Lanzetti, L., Rybin, V., Malabarba, M. G., Christoforidis, S., Scita, G., Zerial, M., & Di Fiore, P. P. 2000, "The Eps8 protein coordinates EGF receptor signalling through Rac and trafficking through Rab5", *Nature*, vol. 408, no. 6810, pp. 374-377.
- Lassle, M., Blatch, G. L., Kundra, V., Takatori, T., & Zetter, B. R. 1997, "Stress-inducible, Murine Protein mSTI1. CHARACTERIZATION OF BINDING DOMAINS FOR HEAT SHOCK PROTEINS AND IN VITRO PHOSPHORYLATION BY DIFFERENT KINASES", *J.Biol.Chem.*, vol. 272, no. 3, pp. 1876-1884.
- Leustek, T., mir-Shapira, D., Toledo, H., Brot, N., & Weissbach, H. 1992, "Autophosphorylation of 70 kDa heat shock proteins", *Cell Mol.Biol.*, vol. 38, no. 1, pp. 1-10.
- Liaw, Y. S., Yang, P. C., Yu, C. J., Kuo, S. H., Luh, K. T., Lin, Y. J., & Wu, M. L. 1998, "PKC activation is required by EGF-stimulated Na(+)-H+ exchanger in human pleural mesothelial cells", *Am.J.Physiol.*, vol. 274, no. 5, Part 1, p. L665-L672.
- Liberatori, S., Canas, B., Tani, C., Bini, L., Buonocore, G., Godovac-Zimmermann, J., Mishra, O. P., ivoria-Papadopoulos, M., Bracci, R., & Pallini, V. 2004, "Proteomic approach to the identification of voltage-dependent anion channel protein isoforms in guinea pig brain synaptosomes", *Proteomics*, vol. 4, no. 5, pp. 1335-1340.
- Lill, J. 2003, "Proteomic tools for quantitation by mass spectrometry", *Mass Spectrom.Rev.*, vol. 22, no. 3, pp. 182-94.
- Link, A. J. 2002, "Multidimensional peptide separations in proteomics", *Trends Biotechnol.*, vol. 20, no. 12 Suppl, pp. 8-13.
- Liu, T., Qian, W. J., Chen, W. N., Jacobs, J. M., Moore, R. J., Anderson, D. J., Gritsenko, M. A., Monroe, M. E., Thrall, B. D., Camp, D. G. 2., & Smith, R. D. 2005, "Improved proteome coverage by using high efficiency cysteinyl peptide enrichment: The human mammary epithelial cell proteome", *Proteomics*, vol. 5, no. 5, pp. 1263-1273.
- Longenecker, K., Read, P., Derewenda, U., Dauter, Z., Liu, X., Garrard, S., Walker, L., Somlyo, A. V., Nakamoto, R. K., Somlyo, A. P., & Derewenda, Z. S. 1999, "How RhoGDI binds Rho", *Acta Crystallogr.D.Biol.Crystallogr.*, vol. 55 ( Pt 9), pp. 1503-1515.
- Lowenstein, E. J., Daly, R. J., Batzer, A. G., Li, W., Margolis, B., Lammers, R., Ullrich, A., Skolnik, E. Y., Bar-Sagi, D., & Schlessinger et, a. 1992, "The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling", *Cell*, vol. 70, no. 3, pp. 431-442.

- Lu, J., McKinsey, T. A., Nicol, R. L., & Olson, E. N. 2000, "Signal-dependent activation of the MEF2 transcription factor by dissociation from histone deacetylases", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 97, no. 8, pp. 4070-4075.
- Maddala, R., Reddy, V. N., Epstein, D. L., & Rao, V. 2003, "Growth factor induced activation of Rho and Rac GTPases and actin cytoskeletal reorganization in human lens epithelial cells", *Mol.Vis.*, vol. 9, pp. 329-336.
- Manevich, Y., Feinstein, S. I., & Fisher, A. B. 2004, "Activation of the antioxidant enzyme 1-CYS peroxiredoxin requires glutathionylation mediated by heterodimerization with {pi}GST", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 101, no. 11, pp. 3780-3785.
- Mann, M., Hendrickson, R. C., & Pandey, A. 2001, "Analysis of proteins and proteomes by mass spectrometry", *Annu.Rev.Biochem.*, vol. 70, no. 1, pp. 437-473.
- Mann, M., Hojrup, P., & Roepstorff, P. 1993, "Use of mass spectrometric molecular weight information to identify proteins in sequence databases", *Biol.Mass Spectrom.*, vol. 22, no. 6, pp. 338-345.
- Mann, M. & Wilm, M. 1994, "Error-tolerant identification of peptides in sequence databases by peptide sequence tags", *Anal.Chem.*, vol. 66, no. 24, pp. 4390-9.
- Mao, Z., Bonni, A., Xia, F., Nadal-Vicens, M., & Greenberg, M. E. 1999, "Neuronal Activity-Dependent Cell Survival Mediated by Transcription Factor MEF2", *Science*, vol. 286, no. 5440, pp. 785-790.
- Marais, R., Light, Y., Mason, C., Paterson, H., Olson, M. F., & Marshall, C. J. 1998, "Requirement of Ras-GTP-Raf complexes for activation of Raf-1 by protein kinase C", *Science*, vol. 280, no. 5360, pp. 109-112.
- Marinissen, M. J., Chiariello, M., & Gutkind, J. S. 2001, "Regulation of gene expression by the small GTPase Rho through the ERK6 (p38{gamma}) MAP kinase pathway", *Genes Dev.*, vol. 15, no. 5, pp. 535-553.
- Marshall, C. J. 1995, "Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation", *Cell*, vol. 80, no. 2, pp. 179-185.
- Martinez, O. & Goud, B. 1998, "Rab proteins", *Biochim.Biophys.Acta*, vol. 1404, no. 1-2, pp. 101-112.
- Mason, D. E. & Liebler, D. C. 2003, "Quantitative analysis of modified proteins by LC-MS/MS of peptides labeled with phenyl isocyanate", *J.Proteome Res.*, vol. 2, no. 3, pp. 265-72.
- Matter, N., Herrlich, P., & Konig, H. 2002, "Signal-dependent regulation of splicing via phosphorylation of Sam68", *Nature*, vol. 420, no. 6916, pp. 691-695.



Mayer, M. P., Brehmer, D., Gassler, C. S., & Bukau, B. 2001, "Hsp70 chaperone machines", *Adv. Protein Chem.*, vol. 59, pp. 1-44.

Mehta, D., Rahman, A., & Malik, A. B. 2001, "Protein Kinase C- $\alpha$  Signals Rho-Guanine Nucleotide Dissociation Inhibitor Phosphorylation and Rho Activation and Regulates the Endothelial Cell Barrier Function", *J. Biol. Chem.*, vol. 276, no. 25, pp. 22614-22620.

Meisner, H. & Czech, M. P. 1995, "Coupling of the Proto-oncogene Product c-Cbl to the Epidermal Growth Factor Receptor", *J. Biol. Chem.*, vol. 270, no. 43, pp. 25332-25335.

Mendelsohn, J. & Baselga, J. 2000, "The EGF receptor family as targets for cancer therapy", *Oncogene*, vol. 19, no. 56, pp. 6550-6565.

Muszynska, G., Andersson, L., & Porath, J. 1986, "Selective adsorption of phosphoproteins on gel-immobilized ferric chelate", *Biochemistry*, vol. 25, no. 22, pp. 6850-3.

Muszynska, G., Dobrowolska, G., Medin, A., Ekman, P., & Porath, J. O. 1992, "Model studies on iron(III) ion affinity chromatography. II. Interaction of immobilized iron(III) ions with phosphorylated amino acids, peptides and proteins", *J. Chromatogr.*, vol. 604, no. 1, pp. 19-28.

Nakashima, I., Kato, M., Akhand, A. A., Suzuki, H., Takeda, K., Hossain, K., & Kawamoto, Y. 2002, "Redox-linked signal transduction pathways for protein tyrosine kinase activation", *Antioxid. Redox. Signal.*, vol. 4, no. 3, pp. 517-531.

Neubauer, G. & Mann, M. 1999, "Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations", *Anal. Chem.*, vol. 71, no. 1, pp. 235-242.

Nollen, E. A. A., Kabakov, A. E., Brunsting, J. F., Kanon, B., Hohfeld, J., & Kampinga, H. H. 2001, "Modulation of in Vivo HSP70 Chaperone Activity by Hip and Bag-1", *J. Biol. Chem.*, vol. 276, no. 7, pp. 4677-4682.

Nuhse, T. S., Stensballe, A., Jensen, O. N., & Peck, S. C. 2003, "Large-scale Analysis of in Vivo Phosphorylated Membrane Proteins by Immobilized Metal Ion Affinity Chromatography and Mass Spectrometry", *Mol. Cell Proteomics*, vol. 2, no. 11, pp. 1234-43.

Oda, Y., Huang, K., Cross, F. R., Cowburn, D., & Chait, B. T. 1999, "Accurate quantitation of protein expression and site-specific phosphorylation", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, no. 12, pp. 6591-6.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J. H., Saito, K., Sakamoto, A., Inoue, M., & Shirouzu et, a. 2002, "Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains", *Cell*, vol. 110, no. 6, pp. 775-787.

Ohkusu-Tsukada, K., Tominaga, N., Udono, H., & Yui, K. 2004, "Regulation of the Maintenance of Peripheral T-Cell Anergy by TAB1-Mediated p38{alpha} Activation", *Mol.Cell Proteomics*, vol. 24, no. 16, pp. 6957-6966.

Olayioye, M. A., Beuvink, I., Horsch, K., Daly, J. M., & Hynes, N. E. 1999, "ErbB receptor-induced activation of stat transcription factors is mediated by Src tyrosine kinases", *J.Biol.Chem.*, vol. 274, no. 24, pp. 17209-17218.

Olofsson, B. 1999, "Rho Guanine Dissociation Inhibitors: Pivotal Molecules in Cellular Signalling", *Cell.Signal.*, vol. 11, no. 8, pp. 545-554.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. 2002, "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics", *Mol.Cell Proteomics*, vol. 1, no. 5, pp. 376-86.

Ong, S. E., Kratchmarova, I., & Mann, M. 2003, "Properties of <sup>13</sup>C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC)", *J.Proteome Res.*, vol. 2, no. 2, pp. 173-181.

Papac, D. I., Oatis, J. E., Jr., Crouch, R. K., & Knapp, D. R. 1993, "Mass spectrometric identification of phosphorylation sites in bleached bovine rhodopsin", *Biochemistry*, vol. 32, no. 23, pp. 5930-5934.

Pappin, D. J., Hojrup, P., & Bleasby, A. J. 1993, "Rapid identification of proteins by peptide-mass fingerprinting", *Curr.Biol.*, vol. 3, no. 6, pp. 327-332.

Pasa-Tolic, J., Jensen, P. K., Anderson, G. A., Lipton, M. S., Peden, K. K., Martinovic, S., Tolic, N., Bruce, J. E., & Smith, R. D. 1999, "High throughput proteome-wide precision measurements of protein expression using mass spectrometry", *J.Am.Chem.Soc.*, vol. 121, no. 34, pp. 7949-7950.

Patton, W. F. 2002, "Detection technologies in proteome analysis", *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.*, vol. 771, no. 1-2, pp. 3-31.

Pawson, T. & Nash, P. 2003, "Assembly of Cell Regulatory Systems Through Protein Interaction Domains", *Science*, vol. 300, no. 5618, pp. 445-452.

Pawson, T. & Scott, J. D. 1997, "Signaling Through Scaffold, Anchoring, and Adaptor Proteins", *Science*, vol. 278, no. 5346, pp. 2075-2080.

Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. 1999, "Probability-based protein identification by searching sequence databases using mass spectrometry data", *Electrophoresis*, vol. 20, no. 18, pp. 3551-67.

Porath, J. 1990, "Amino acid side chain interaction with chelate-liganded crosslinked dextran, agarose and TSK gel. A mini review of recent work", *J.Mol.Recognit.*, vol. 3, no. 3, pp. 123-7.

- Posewitz, M. C. & Tempst, P. 1999, "Immobilized gallium(III) affinity chromatography of phosphopeptides", *Anal.Chem*, vol. 71, no. 14, pp. 2883-2892.
- Prenzel, N., Zwick, E., Daub, H., Leserer, M., Abraham, R., Wallasch, C., & Ullrich, A. 1999, "EGF receptor transactivation by G-protein-coupled receptors requires metalloproteinase cleavage of proHB-EGF", *Nature*, vol. 402, no. 6764, pp. 884-888.
- Quemeneur, E., Guthapfel, R., & Gueguen, P. 1994, "A major phosphoprotein of the endoplasmic reticulum is protein disulfide isomerase", *J.Biol.Chem.*, vol. 269, no. 8, pp. 5485-5488.
- Rak, A., Pylypenko, O., Durek, T., Watzke, A., Kushnir, S., Brunsfeld, L., Waldmann, H., Goody, R. S., & Alexandrov, K. 2003, "Structure of Rab GDP-Dissociation Inhibitor in Complex with Prenylated YPT1 GTPase", *Science*, vol. 302, no. 5645, pp. 646-650.
- Reid, G. E. & McLuckey, S. A. 2002, "'Top down' protein characterization via tandem mass spectrometry", *J Mass Spectrom*, vol. 37, no. 7, pp. 663-75.
- Reiss, N., Kanety, H., & Schlessinger, J. 1986, "Five enzymes of the glycolytic pathway serve as substrates for purified epidermal-growth-factor-receptor kinase", *Biochem.J.*, vol. 239, no. 3, pp. 691-697.
- Ren, Y., Meng, S., Mei, L., Zhao, Z. J., Jove, R., & Wu, J. 2003, "Roles of Gab1 and SHP2 in paxillin tyrosine dephosphorylation and Src activation in response to epidermal growth factor", *J.Biol.Chem.*
- Roos, M., Soskic, V., Poznanovic, S., & Godovac-Zimmermann J. 1998, "Post-translational Modifications of Endothelin Receptor B from Bovine Lungs Analyzed by Mass Spectrometry", *J.Biol.Chem.*, vol. 273, no. 2, pp. 924-931.
- Sakaguchi, K., Okabayashi, Y., Kido, Y., Kimura, S., Matsumura, Y., Inushima, K., & Kasuga, M. 1998, "Shc phosphotyrosine-binding domain dominantly interacts with epidermal growth factor receptors and mediates Ras activation in intact cells", *Mol.Endocrinol.*, vol. 12, no. 4, pp. 536-543.
- Santos, M. F., McCormack, S. A., Guo, Z., Okolicany, J., Zheng, Y., Johnson, L. R., & Tigyi, G. 1997, "Rho Proteins Play a Critical Role in Cell Migration during the Early Phase of Mucosal Restitution", *J.Clin.Invest.*, vol. 100, no. 1, pp. 216-225.
- Sasaoka, T., Langlois, W. J., Leitner, J. W., Draznin, B., & Olefsky, J. M. 1994, "The signaling pathway coupling epidermal growth factor receptors to activation of p21ras", *J.Biol.Chem.*, vol. 269, no. 51, pp. 32621-32625.
- Sato, S., Fujita, N., & Tsuruo, T. 2000, "Modulation of Akt kinase activity by binding to Hsp90", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 97, no. 20, pp. 10832-10837.

- Scheffzek, K., Stephan, I., Jensen, O. N., Illenberger, D., & Gierschik, P. 2000, "The Rac-RhoGDI complex and the structural basis for the regulation of Rho proteins by RhoGDI", *Nat.Struct.Biol.*, vol. 7, no. 2, pp. 122-126.
- Schlessinger, J. 2000, "Cell signaling by receptor tyrosine kinases", *Cell*, vol. 103, no. 2, pp. 211-225.
- Schoenmakers, C. H., Pigmans, I. G., Hawkins, H. C., Freedman, R. B., & Visser, T. J. 1989, "Rat liver type I iodothyronine deiodinase is not identical to protein disulfide isomerase", *Biochem.Biophys.Res.Comm.*, vol. 162, no. 2, pp. 857-868.
- Schulenberg, B., Goodman, T. N., Aggeler, R., Capaldi, R. A., & Patton, W. F. 2004, "Characterization of dynamic and steady-state protein phosphorylation using a fluorescent phosphoprotein gel stain and mass spectrometry", *Electrophoresis*, vol. 25, no. 15, pp. 2526-2532.
- Seabra, M. C. & Wasmeier, C. 2004, "Controlling the location and activation of Rab GTPases", *Curr.Opin.Cell Biol.*, vol. 16, no. 4, pp. 451-457.
- Shen, Y., Tolic, N., Masselon, C., Pasa-Tolic, L., Camp, D. G. 2., Hixson, K. K., Zhao, R., Anderson, G. A., & Smith, R. D. 2004, "Ultrasensitive proteomics using high-efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS", *Anal.Chem.*, vol. 76, no. 1, pp. 144-54.
- Shen, Y., Meunier, L., & Hendershot, L. M. 2002, "Identification and Characterization of a Novel Endoplasmic Reticulum (ER) DnaJ Homologue, Which Stimulates ATPase Activity of BiP in Vitro and Is Induced by ER Stress", *J.Biol.Chem.*, vol. 277, no. 18, pp. 15947-15956.
- Sherman, M. Y. & Goldberg, A. L. 1993, "Heat Shock of Escherichia coli Increases Binding of dnaK (the hsp70 Homolog) to Polypeptides by Promoting its Phosphorylation", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 90, no. 18, pp. 8648-8652.
- Sherrill, J. M. 1997, "Insufficiency of self-phosphorylation for the activation of epidermal growth factor receptor", *Biochemistry*, vol. 36, no. 19, pp. 5677-5684.
- Shevchenko, A., Wilm, M., Vorm, O., & Mann, M. 1996, "Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels", *Anal.Chem.*, vol. 68, no. 5, pp. 850-858.
- Shi, S. H., Jan, L. Y., & Jan, Y. N. 2003, "Hippocampal neuronal polarity specified by spatially localized mPar3/mPar6 and PI 3-kinase activity", *Cell*, vol. 112, no. 1, pp. 63-75.
- Shisheva, A., Buxton, J., & Czech, M. P. 1994, "Differential intracellular localizations of GDP dissociation inhibitor isoforms. Insulin-dependent redistribution of GDP dissociation inhibitor-2 in 3T3-L1 adipocytes", *J.Biol.Chem.*, vol. 269, no. 39, pp. 23865-23868.
- Sitia, R. & Molteni, S. N. 2004, "Stress, Protein (Mis)folding, and Signaling: The Redox Connection", *Sci.STKE*, vol. 2004, no. 239, p. e27.

- Smart, E. J., Graf, G. A., McNiven, M. A., Sessa, W. C., Engelman, J. A., Scherer, P. E., Okamoto, T., & Lisanti, M. P. 1999, "Caveolins, liquid-ordered domains, and signal transduction", *Mol. Cell Biol.*, vol. 19, no. 11, pp. 7289-7304.
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., & Klenk, D. C. 1985, "Measurement of protein using bicinchoninic acid", *Anal. Biochem.*, vol. 150, no. 1, pp. 76-85.
- Song, J., Takeda, M., & Morimoto, R. I. 2001, "Bag1-Hsp70 mediates a physiological stress signalling pathway that regulates Raf-1/ERK and cell growth", *Nat. Cell Biol.*, vol. 3, no. 3, pp. 276-282.
- Sorkina, T., Bild, A., Tebar, F., & Sorkin, A. 1999, "Clathrin, adaptors and eps15 in endosomes containing activated epidermal growth factor receptors", *J. Cell Sci.*, vol. 112, no. 3, pp. 317-327.
- Sorokin, A., Lemmon, M. A., Ullrich, A., & Schlessinger, J. 1994, "Stabilization of an active dimeric form of the epidermal growth factor receptor by introduction of an inter-receptor disulfide bond", *J. Biol. Chem.*, vol. 269, no. 13, pp. 9752-9759.
- Soskic, V., Gorlach, M., Poznanovic, S., Boehmer, F. D., & Godovac-Zimmermann, J. 1999, "Functional proteomics analysis of signal transduction pathways of the platelet-derived growth factor beta receptor", *Biochemistry*, vol. 38, no. 6, pp. 1757-64.
- Stannard, C., Brown, L. R., & Godovac-Zimmermann J. 2004, "New paradigms in cellular function and the need for top-down proteomics analysis", *Current Proteomics*, vol. 1, pp. 13-25.
- Stannard, C., Lehenkari, P., & Godovac-Zimmermann, J. 2003a, "Functional diversity of endothelin pathways in human lung fibroblasts may be based on structural diversity of the endothelin receptors", *Biochemistry*, vol. 42, no. 47, pp. 13909-13918.
- Stannard, C., Soskic, V., & Godovac-Zimmermann, J. 2003b, "Rapid changes in the phosphoproteome show diverse cellular responses following stimulation of human lung fibroblasts with endothelin-1", *Biochemistry*, vol. 42, no. 47, pp. 13919-13928.
- Steen, H., Kuster, B., Fernandez, M., Pandey, A., & Mann, M. 2002, "Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway", *J. Biol. Chem.*, vol. 277, no. 2, pp. 1031-9.
- Sun, L. & Carpenter, G. 1998, "Epidermal growth factor activation of NF-kappaB is mediated through IkappaBalpha degradation and intracellular free calcium", *Oncogene*, vol. 16, no. 16, pp. 2095-2102.
- Sze, S. K., Ge, Y., Oh, H., & McLafferty, F. W. 2002, "Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 4, pp. 1774-9.

Takayama, S. & Reed, J. C. 2001, "Molecular chaperone targeting and regulation by BAG family proteins", *Nat. Cell Biol.*, vol. 3, no. 10, p. E237-E241.

Tanno, M., Bassi, R., Gorog, D. A., Saurin, A. T., Jiang, J., Heads, R. J., Martin, J. L., Davis, R. J., Flavell, R. A., & Marber, M. S. 2003, "Diverse Mechanisms of Myocardial p38 Mitogen-Activated Protein Kinase Activation: Evidence for MKK-Independent Activation by a TAB1-Associated Mechanism Contributing to Injury During Myocardial Ischemia", *Circ. Res.*, vol. 93, no. 3, pp. 254-261.

Tice, D. A., Biscardi, J. S., Nickles, A. L., & Parsons, S. J. 1999, "Mechanism of biological synergy between cellular Src and epidermal growth factor receptor", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, no. 4, pp. 1415-1420.

Topp, J. D., Gray, N. W., Gerard, R. D., & Horazdovsky, B. F. 2004, "Alsin Is a Rab5 and Rac1 Guanine Nucleotide Exchange Factor", *J. Biol. Chem.*, vol. 279, no. 23, pp. 24612-24623.

Townsend, P. A., Cutress, R. I., Sharp, A., Brimmell, M., & Packham, G. 2003, "BAG-1: a multifunctional regulator of cell growth and survival", *Biochim. Biophys. Acta-Rev. Cancer*, vol. 1603, no. 2, pp. 83-98.

Trauger, S. A., Go, E. P., Shen, Z., Apon, J. V., Compton, B. J., Bouvier, E. S., Finn, M. G., & Siuzdak, G. 2004, "High Sensitivity and Analyte Capture with Desorption/Ionization Mass Spectrometry on Silylated Porous Silicon", *Anal. Chem.*, vol. 76, no. 15, pp. 4484-4489.

Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., & Schlessinger et, a. 1984, "Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells", *Nature*, vol. 309, no. 5967, pp. 418-425.

van der Spuy, J., Cheetham, M. E., Dirr, H. W., & Blatch, G. L. 2001, "The Cochaperone Murine Stress-Inducible Protein 1: Overexpression, Purification, and Characterization", *Protein Expr. Purif.*, vol. 21, no. 3, pp. 462-469.

Velten, M., Villoutreix, B. O., & Ladjimi, M. M. 2000, "Quaternary structure of the HSC70 cochaperone HIP", *Biochemistry*, vol. 39, no. 2, pp. 307-315.

Vendemiale, G., Guerrieri, F., Grattagliano, I., Didonna, D., Muolo, L., & Altomare, E. 1995, "Mitochondrial oxidative phosphorylation and intracellular glutathione compartmentation during rat liver regeneration", *Hepatology*, vol. 21, no. 5, pp. 1450-1454.

Wang, H. & Hanash, S. 2004, "Intact-protein based sample preparation strategies for proteome analysis in combination with mass spectrometry", *Mass Spectrom. Rev.*, vol. 30, p. 30.

Wang, H. G., Rapp, U. R., & Reed, J. C. 1996, "Bcl-2 targets the protein kinase Raf-1 to mitochondria", *Cell*, vol. 87, no. 4, pp. 629-638.

Ware, M. F., Tice, D. A., Parsons, S. J., & Lauffenburger, D. A. 1997, "Overexpression of Cellular Src in Fibroblasts Enhances Endocytic Internalization of Epidermal Growth Factor Receptor", *J.Biol.Chem.*, vol. 272, no. 48, pp. 30185-30190.

Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M., & Yates, J. R. 3. 2002, "Analysis of quantitative proteomic data generated via multidimensional protein identification technology", *Anal.Chem.*, vol. 74, no. 7, pp. 1650-7.

Weiller, G. F., Caraux, G., & Sylvester, N. 2004, "The modal distribution of protein isoelectric points reflects amino acid properties rather than sequence evolution", *Proteomics*, vol. 4, no. 4, pp. 943-949.

Xia, L., Wang, L., Chung, A. S., Ivanov, S. S., Ling, M. Y., Dragoi, A. M., Platt, A., Gilmer, T. M., Fu, X. Y., & Chin et, a. 2002, "Identification of both positive and negative domains within the epidermal growth factor receptor COOH-terminal region for signal transducer and activator of transcription (STAT) activation", *J.Biol.Chem.*, vol. 277, no. 34, pp. 30716-30723.

Yaffe, M. B. & Elia, A. E. H. 2001, "Phosphoserine/threonine-binding domains", *Curr.Opin.Cell Biol.*, vol. 13, no. 2, pp. 131-138.

Yamauchi, T., Ueki, K., Tobe, K., Tamemoto, H., Sekine, N., Wada, M., Honjo, M., Takahashi, M., Takahashi, T., & Hirai et, a. 1997, "Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone", *Nature*, vol. 390, no. 6655, pp. 91-96.

Yates, J. R. 3. 2004, "Mass spectral analysis in proteomics", *Annu.Rev.Biophys.Biomol.Struct.*, vol. 33, pp. 297-316.

Yoshimura, T., Kawano, Y., Arimura, N., Kawabata, S., Kikuchi, A., & Kaibuchi, K. 2005, "GSK-3[beta] Regulates Phosphorylation of CRMP-2 and Neuronal Polarity", *Cell*, vol. 120, no. 1, pp. 137-149.

Zachariou, M. & Hearn, M. T. 2000, "Adsorption and selectivity characteristics of several human serum proteins with immobilised hard Lewis metal ion-chelate adsorbents", *J.Chromatogr.A*, vol. 890, no. 1, pp. 95-116.

Zhou, W., Merrick, B. A., Khaledi, M. G., & Tomer, K. B. 2000, "Detection and sequencing of phosphopeptides affinity bound to immobilized metal ion beads by matrix-assisted laser desorption/ionization mass spectrometry", *J.Am.Soc.Mass Spectrom.*, vol. 11, no. 4, pp. 273-282.

Zorzano, A., Bach, D., Pich, S., & Palacin, M. 2004, "Role of novel mitochondrial proteins in energy balance", *Rev.Med.Univ.Navarra*, vol. 48, no. 2, pp. 30-35.